NAVAL RESEARCH LOGISTICS QUARTERLY

OFFICE OF NAVAL RESEARCH

CONTENTS

ARTICLES	Page
Improvise or Plan? *CAPT A. H. Wilson and CDR W. R. Finn*	263
Some Production Planning Problems *M. Klein*	269
Simulation Research on Job Shop Production *J. R. Jackson*	287
Mathematical Programming and Employment Scheduling *W. Karush and A. Vazsonyi*	297
Formulation of Recurrence Equations for Shuttle Process and Assembly Line *R. Bellman*	321
Relationships Between Weapons and Logistics Expenditures *M. A. Geisler*	335
An Application of Game Theory to Special Weapons Evaluation *J. K. Hale and H. H. Wicke*	347
An Optimal Search Pattern *J. R. Isbell*	357
NOTES	361
NEWS AND MEMORANDA	363
RECENT PUBLICATIONS	369
AUTHOR INDEX FOR VOLUME IV	371

Los Angeles Public Library
DEPOSITORY
JAN 16 1958

VOL. 4, NO. 4 DECEMBER 1957

NAVEXOS P-1278

NAVAL RESEARCH LOGISTICS QUARTERLY

EDITORS

Rear Admiral H. E. Eccles, USN (Retired)
The George Washington University

F. D. Rigby
Office of Naval Research

O. Morgenstern
Princeton University

Commander M. I. Rosenberg, USN
Managing Editor
Office of Naval Research
Washington 25, D. C.

ASSOCIATE EDITORS

J. S. Campbell, *Office of Naval Research*
E. W. Cannon, *National Bureau of Standards*
T. B. Evans, *Brigadier General, USA*
P. L. Folsom, *Captain, USN*
M. A. Geisler, *The Rand Corporation*
E. B. Henry, Jr., *Commander, USN*
A. J. Hoffman, *General Electric Company*
S. Karlin, *Stanford University*
J. Laderman, *Office of Naval Research Br. Office*
A. T. Magnell, *Captain, SC, USN*
W. H. Marlow, *The George Washington University*
J. Marschak, *Cowles Foundation*
Vice Admiral R. E. McShane, *USN (Retired)*
W. F. Millson, *Captain SC, USN*
H. A. Sachaklian, *Colonel, USAF*
E. K. Scofield, *Captain, SC, USN*
J. S. Skoczylas, *Lt. Colonel, USMC*
E. D. Stanley, *Captain, SC, USN*
C. Stein, Jr., *Captain, SC, USN*
R. M. Thrall, *University of Michigan*
C. B. Tompkins, *University of California*
A. W. Tucker, *Princeton University*
J. D. Wilkes, *ONR*
M. Wood, *Office of the Asst. Sec. of Defense*

The Naval Research Logistics Quarterly is devoted to the dissemination of scientific information in logistics and will publish research and expository papers, including those in certain areas of mathematics, statistics, and economics, relevant to the over-all effort to improve the efficiency and effectiveness of logistics operations.

Information for Contributors is indicated on back cover.

The Naval Research Logistics Quarterly is published by the Office of Naval Research in the months of March, June, September, and December and can be purchased from the Superintendent of Documents, U. S. Government Printing Office, Washington 25, D. C. Subscription Price: $1.50 a year in the U. S. and Canada, $2.00 elsewhere; 50¢ for a single copy. Reprints of an individual article can be purchased in quantities of 100 or more. Requests for the purchase price of reprints of a particular article should be sent to the Superintendent of Documents.

The views and opinions expressed in this Quarterly are those of the authors and not necessarily those of the Office of Naval Research.

The printing of this publication was approved by the Director of the Bureau of the Budget, July 26, 1955

IMPROVISE OR PLAN?

CAPT A. H. Wilson
U. S. Navy

and

CDR W. R. Finn
SC, U. S. Navy

Often our military thinking reflects past experience instead of future concepts. Often we have entered a conflict with the doctrines, organizations, and support systems that represent, at best, an improvement over those employed in the previous war. Then, as the situation develops, and as we learn on the scene, we improvise to meet the requirements that time and technology have cast. We develop new weapons, new weapons systems, and new hardware. We dream and plan our future strategies. Yet, frequently, the doctrines and organizations for the support systems are incompatible with new developments and are, by contrast, archaic and neglected.

During the past few years, much has been written concerning changes in military concepts and doctrines required in the light of new weapons systems. Most publicized, perhaps, has been the Army thinking. The Marine Corps, too, has proclaimed radical departures in their tactics and methods. The Air Force concepts are obvious in their recognition of the nuclear age. What about the Navy?

The Navy has been increasing capabilities constantly. We are acquiring the most modern weapons systems. We are exploiting many means of research. How do these factors influence our organizations—our tactics—and our logistic support doctrines? Are we planning for these also? Or will we improvise when the time comes? The writers have been most interested in the logistic aspects of this situation and believe it is an area that has been most subject to improvisation in the past.

At the risk of over-simplification and in an effort to concentrate on the logistics implications of modern weapons, there will be set forth in this paper some considerations concerning the support of an Attack Carrier Striking Force that may provoke discussion. The aspects of the subject that will be considered are: planning for requirements, Mobile Logistic Support Group operations and doctrine, Underway Replenishment Group operations and doctrine, and other related problems.

At the onset of the discussion, we should immediately examine the concepts voiced by individuals, study groups, and current doctrines for naval forces involved in an Attack Carrier

Striking Force, including its composition, disposition, and employment. Here we find concepts which range from forces of from one to four aircraft carriers in a group; with or without guided missile cruisers; with or without heavy support ships; with or without escorts; and with various additions to and combinations of the foregoing. We find concepts requiring tremendous dispersion (in miles) between units and concepts following modified World War II dispositions. We find hit-and-run concepts and "endurance-on-the-line" concepts. In spite of the broad variances in the foregoing, it is believed we can look at some of the logistic implications involved in some of the more radically changed force concepts.

REQUIREMENTS PLANNING

What will be the impact of nuclear weapons upon requirements planning for logistic support in all categories of supply?

The greatest impact will probably be upon ammunition requirements. The accepted planning factors for estimating ammunition will no longer apply. If nuclear weapons have the effects conceded, the requirements for conventional ammunition for an Attack Carrier Striking Force may be reduced to that required for the air defense of the task force and may soon be displaced by ground-to-air and air-to-air missiles. If we still assume World War II type air attacks against surface units and continue manning and firing obsolete AA batteries, the force has within its own ammunition allowance sufficient defensive ammunition for approximately 100 average World War II enemy aircraft attack engagements.

With nuclear weapons capable of predicted yields, the amount of destruction achieved by one successful air strike will exceed that of an entire Attack Carrier Striking Force launching full-strike capability with conventional weapons. The amount of destruction one Attack Carrier is capable of achieving reduces the number of carriers required and conceivably reduces the size of the Attack Carrier Striking Force. The total effect of the foregoing would be to reduce the ammunition requirements to within the ships' ammunition allowances, with the end result of eliminating the need for ammunition ships in replenishment at sea.

Next, let us consider the corresponding influence upon fuel consumption. If through the use of nuclear weapons the size of the Attack Carrier Striking Force can be markedly reduced, as discussed above, the requirements for Navy special fuel will be reduced accordingly. Furthermore, the length of time required to achieve the level of destruction desired is reduced to a few sorties. For over-simplification, assume that strikes against all targets of the system under attack are launched the same day, with clean-up for aborts within the same 24-hour period. Then, the time the Attack Carrier Striking Force is on the line could be reduced to 24 hours. Hence, all the foregoing could reduce the requirements for navy special fuel to the amount required to steam the force to the objective area and return.

Similarly, the requirements for aviation fuels will be greatly reduced. If the size of a strike group is reduced to one attack aircraft plus escort, the number of strikes reduced to one on each selected target (disregarding aborts and combat attrition), and the over-all protective missions reduced because of reduced time the force is on the line, then the aviation fuel consumption is drastically reduced. In fact, the fuel requirements for an Attack Carrier Striking Force may be reduced to the capacity of the force.

Because the requirements for ammunition and fuel are invariably the governing factors in resupply, the effect of nuclear weapons on other categories of supply will not be discussed, other than to say these requirements would be within the inherent capacity and endurance of the combatant ships of the force.

Some may argue that the nuclear weapons available in the force may not be sufficient to insure the level of destruction desired. Hence, either additional carriers or replenishment of weapons at sea may be required. Further, progressive target destruction with evaluation and analysis of damage may be desirable, rather than an all-out, one-time attack. This would require longer periods on the line, requiring replenishment at sea, at least for fuel.

Finally, some contend that limited target destruction may be desirable against certain installations because of political or economic reasons. Or, they say, there is the possibility that military occupation may dictate use of conventional weapons. Such arguments ignore the flexibility and capability of tactical nuclear weapons.

EFFECT ON MOBILE SUPPORT GROUP OPERATIONS AND DOCTRINE

How would the employment of nuclear weapons effect current doctrine for mobile support group operations?

Consider first the effect of the employment of these weapons by our own forces. If the reduction in requirements, as discussed above, is considered valid, there will no longer be need for heavy concentrations of point-to-point shipping and shuttle units (to the Underway Replenishment Group) at advanced bases. A reduction in bottoms for point-to-point shipping may result in the building of replenishment types sufficient to permit fleet issue loading of all resupply shipping. The reduction of time in the objective area and the reduction in number of ships in the Attack Carrier Striking Force will reduce the maintenance and repair requirement at the advanced bases. And reduction of ammunition and fuel requirements will eliminate the requirement for heavy reserve storage at advanced bases.

All of the foregoing tends toward a reduction in the functions and services performed by the Mobile Support Group. A reduction in functions and services automatically reduces the number of service craft, volume of point-to-point shipping, and personnel. All of which reduces the demand for large harbor and port administrative units, cargo handling battalions, and so on.

Current mobile support doctrine calls for reloading shuttle units of the Underway Replenishment Group at advanced bases. But if the reasoning outlined in the foregoing paragraphs is carried to an extreme and the requirements are reduced to the extent that point-to-point shipping for fleet support is eliminated, then there will be little requirement for the advanced base as we know it today.

Now, let us consider the effect upon current doctrine when nuclear weapons are employed by the enemy. Heavy concentrations of logistic support units and point-to-point shipping at advanced bases which are vulnerable to air attack will no longer be acceptable. As the base situation becomes untenable, current doctrine must be modified. Conceivably some modifications would include the following. Shuttle units would return to rear areas or the Continental United States for reloading, except for fleet oilers which would receive liquid cargos at sea. Repair and maintenance units would no longer be included in the Mobile Support Group, except for towing and salvage units which would remain at sea. Services such as mail, movies, and fleet freight would be reduced to that capable of being transferred between fleet oilers and tankers at sea. Reserves of fuel, ammunition, or other critical items would be retained afloat at secure rear areas.

If we were permitted by the enemy to occupy an advanced base, the concentration of shipping and shore activity would have to be reduced below the level of profitable enemy attack. To accomplish this and to utilize the base would require new organizations, dispersions,

increased transfer rates, reduced turn-around times, revised convoy schedules and practices—and new logistic concepts and procedures.

EFFECT ON UNDERWAY REPLENISHMENT GROUP OPERATIONS AND DOCTRINE

As the Mobile Support Group and the Underway Replenishment Group operations are the opposite sides of the same coin, i.e., mobile logistic support, many of the considerations affecting one side relate to the other. The previous discussion of Mobile Support Group Operations effects current doctrine of Underway Replenishment Group Operations. Let us examine the subject further, however.

In regard to the effect on underway replenishment by our own use of nuclear weapons, we have seen both our requirements of fuel and ammunition reduced. The reduced volume of resupply would permit shuttle units to return to rear areas. Carried to the extreme and with the size of the Attack Carrier Striking Force reduced to a minimum, rotation of combatant ships, particularly carriers, to the task force from rear areas may replace methods of replenishment at sea. The current unsolved problem of furnishing replacement aircraft and of off-loading duds would disappear. There may remain a requirement for sending up a single high-speed oiler to top off the force and then depart. Or the requirement for a one-stop weapon and fuel replenishment ship may arise.

With the employment of nuclear weapons by the enemy, the effect upon our current doctrine of underway replenishment would be far-reaching. This is reflected in the radically changed concept, as outlined above, with obvious drastic changes in doctrine and operations of replenishment at sea. Consider for a moment, however, the situation of a limited or conventional war where nuclear weapons are not employed at the onset but where the threat and possibility of the enemy using them remains constant. Under such conditions, the requirements of resupply are not reduced, and replenishment at sea requirements remain at current levels. The concentration of the Attack Carrier Striking Force and Underway Replenishment Group vessels would be such as to offer profitable targets for nuclear weapon attack. The underway replenishment operations would be conducted at the extremes of enemy aircraft range—an almost impossible situation and an operational dilemma. The Attack Carrier Striking Force may be reduced to smaller, widely dispersed, formations. Large, high-speed, single units of radical design, capable of replenishing all categories, may be incorporated into the Attack Carrier Striking Force and rotated for replenishment purposes. Reduction of Underway Replenishment Groups to small or single units would radically change escort requirements. Air protection for replenishment units would pose other problems.

RELATED PROBLEMS

Once the probable future requirements, support concepts, and doctrines are determined, what are some of the related problems?

As all requirements are reduced, diminished pipeline requirements may be anticipated. Changes in replenishment ship designs and modifications to existing vessels also have been indicated. The over-all reduction of escort vessels for the Attack Carrier Striking Force and Mobile Logistic Support Force may permit utilization of these types in other tasks. What then is the fleet logistic problem in support of other naval tasks? How do these compare with problems of the Attack Carrier Striking Force: What is the balance problem? What is the effect on personnel, rating structures, allocations, and training?

SUMMARY AND CONCLUSIONS

The foregoing discussion may indicate a decrease in the quantity of fleet logistic support, especially in the objective area. But this same discussion should indicate a need for an increase in the quality of the support, as well as for increased preciseness and timing. It follows that such needs require both planning and study. Imagination and boldness of thinking challenge logistic planners. To improvise is to abet the enemy.

Finally, how does all this effect the Zone of the Interior logistic effort? Lest we be misled, history discloses that technology increases, not diminishes, the logistic problem. Where do we look for the joker in the stacked deck? It is believed the resulting requirements for dispersal, alternative facilities, skilled manpower, and communications in the rear area and in the Continental United States will pose some fantastic, expensive, and perplexing problems and dilemmas. As in the fleet support concepts, so in the Zone of the Interior support concepts and organizations the problem of the support of naval forces in the face of nuclear weapons requires real thought, study, and planning. To wait and improvise may result in too little, too late—too bad.

* * *

SOME PRODUCTION PLANNING PROBLEMS[1]

Morton Klein*
Columbia University

I. INTRODUCTION

Two closely related groups of production planning problems will be examined. They are (1) allocation problems and (2) scheduling problems.

As examples of interest in the first group, consider the following:

The production cut-back scheduling problem: A plant which has been operating at full capacity is faced with an interval of low demand; a production cut-back is decided upon. A further decision to be made is to reschedule production. This is to be accomplished so as to minimize the total cost of production.

A rationing problem: In a time of raw-material shortage, a multiproduct firm, using the same base for all of its products, has to decide how to allocate its ration among the different products so as to maximize its profits. There will usually be constraints on the quantities which can be used by each product line.

Process mix problem: Several machines are available for the production of a specified quantity of parts. The available time for production on each machine is limited. The problem is to allocate the production among the machines in such a way as to minimize costs.

Economic lot size problem: A firm manufactures several products. If setup charges and inventory holding costs are different for each product, and if there is a constraint on the total quantity of goods which can be stored, what is the economic lot size for each product? Additional restrictions in the form of limited storage space available for particular products are assumed.

It can be seen that these and other similarly structured problems may be given an encompassing mathematical expression. It will be a convenience to discuss this class of problems in the notation and language appropriate to the production scheduling problem given below.

The basic scheduling problem: Given a fixed quantity of material to produce on or before a specified delivery date, production is to be scheduled among the number of periods available in such a way as to minimize total costs. It is assumed that the capacity of the plant or machine being utilized is known for each period within the interval of interest. Capacity

*Manuscript received April 8, 1957
[1]This paper is based on a dissertation submitted in partial fulfillment of the requirements for the degree of Doctor of Engineering Science, in the School of Engineering, Columbia University

here is defined to be the maximum output attainable within the structure of applicable restrictions imposed on the plant. These may be mechanical, social, temporal, spatial, environmental, etc.

The second group of problems is not as large as the first. It is typified by those described below.

The scheduling problem for a series of requirements: The problem statement and assumptions are the same as those of the basic scheduling problem, except that instead of scheduling production to meet a terminal demand only, an intermediary series of requirements in the form of a minimal cumulative schedule at different periods also must be fulfilled.

A purchasing problem: A given amount of raw material has to be purchased and made available for production use according to a given schedule. There are limits on the capital and/or storage space available, and there are additional costs involved in carrying the inventory as a function of the quantity held and the length of time in storage. The problem is to determine the purchasing schedule which minimizes the sum of all costs involved.

Recent papers by Modigliani and Hohn [7] and Charnes, Cooper, and Mellon [2] have produced important insights into the type of information required for the solution of the problem of scheduling to meet a series of requirements under certain conditions. Modigliani and Hohn restricted their attention to the case in which the cumulative production cost function was the same for each period in the total time interval of interest and was a monotone increasing differentiable convex function of output. For this kind of cost function, they demonstrated that the complete set of future requirements need not be known in order to plan production in an optimal fashion. Instead, demand information at only some isolated critical periods is needed for making scheduling decisions. The critical periods are those in which the total problem decomposes into independent subproblems within which intermediate scheduling requirements do not affect the results. Further, they showed that if inventory costs are negligible, under the above costs conditions the optimal production plan was to divide the required quantity evenly among each group of periods within a subproblem. Of even greater interest, however, they pointed out that the precise parametric form of the cost function had no effect on the solution of this problem. Thus, in order to apply their results, information relative only to the convexity property of the cumulative production cost function need be obtained. The far-reaching consequences of these results need not be detailed.

Charnes, Cooper, and Mellon extended the Modigliani and Hohn results to the slightly more general condition of a continuous, nondecreasing, convex cost function. They dropped the requirement of differentiability. They also pointed out that more than one product and factor could be incorporated into the model, provided that suitable surrogates for total costs be utilized instead of output. In particular, they advocated the appropriateness of labor hours of input as a substitute for output for the domain of the combined cost function. Their interest was confined to production costs; hence, they omitted inventory considerations in their model.

There has been some work on production scheduling problems in which a more general model was used than that considered here. For example, under the assumption that all costs are linear functions, Johnson and Dantzig [6] have considered, in addition to production and inventory costs, the costs of changing the rate of production. Their computational technique is based on the well-known duality theorem of linear programming and encompasses earlier work by Hoffman and Jacobs [3] and Antosiewicz and Hoffman [1] on linear models. Also, the work of Holt, Modigliani, Simon, and Muth [4,5] utilizes a special model which assume that costs are quadratic functions of output and changes in employment level.

II. COST FUNCTIONS

Our attention will be confined to allocation and scheduling problems involving costs which are either convex or concave functions defined over the domains appropriate for each problem.

The following are representative of the class of concave functions:

1. $f(x) = c + bx$
2. $f(x) = \log x \qquad x > 0$

Typical convex functions are:

1. $f(x) = c + bx$
2. $f(x) = x^c \qquad c \geq 1,\ x \geq 0$

More precisely, if x_1 and x_2 are two possible values of the independent variable, and $0 \leq a \leq 1$, then one of the following conditions holds true:

(1.1) Concavity: $f_i(ax_1 + (1-a)\, x_2) \geq af_i(x_1) + (1-a)\, f_i(x_2)$

(1.2) Convexity: $f_i(ax_1 + (1-a)\, x_2) \leq af_i(x_1) + (1-a)\, f_i(x_2)$

III. THE BASIC SCHEDULING PROBLEM

As was previously suggested, the class of allocation problems of interest here may be formulated and discussed in terms of the basic scheduling problem. The translations required in order to have the formulation given below coherent for the other problems is quite clear.

Let: S denote the total quantity of goods required to be available by the end of the n^{th} period,

n denote the total number of periods (days, weeks, months) within the interval of time spanned by the problem,

x_i be the amount to be produced during the i^{th} period,

$f_i(x_i)$ be the total cost associated with producing the quantity x_i,

c_i denote the productive capacity of the system during the i^{th} period, and

$q(x_1, x_2, \ldots, x_n)$ denote the total cost associated with the choice of a production plan, $(x_1, x_2, \ldots, x_n)$.

The basic scheduling problem is to choose a production plan, $(x_1, x_2, \ldots x_n)$, which minimizes the total costs,

(1.3) $$q(x_1, x_2, \ldots, x_n) = \sum f_i(x_i)$$

subject to the constraints

$$\sum x_i = S \text{ , where } 0 \leq S \leq \sum c_i \text{ ,} \tag{1.4}$$

$$0 \leq x_i \leq c_i \text{ ,} \qquad i = 1, 2, \ldots, n \tag{1.5}$$

and where either (1.1) or (1.2) holds for all cost functions, f_i ($i = 1, 2, \ldots, n$).

The applicability of the form of the expressions for total costs given by (1.3) is easily demonstrated.

The most elementary situation occurs if it is assumed that the cost of production as a function of output is invariant with time. This was the assumption of Modigliani and Hohn. Further, it is conventional to assume that inventory costs may be expressed as a linear function of the time that a unit is in storage. Hence, the total cost may be written as

$$q = \sum \left[h(x_i) + b(n-i)\, x_i \right], \tag{1.6}$$

where b = the unit inventory cost per unit of time

$h(x_i)$ = the cost of production as a function of output.

Let

$$f_i(x_i) = h(x_i) + b(n-i)\, x_i \text{ ;}$$

then the cost function, q, takes the form given by (1.3).

Further, if it is assumed that $h(x_i)$ is a concave (convex) function, then since $b\,(n-i)x_i$ is a linear function, it follows that $f_i(x_i)$, ($i = 1, 2 \ldots n$) are all concave (convex) functions.

It may be noted that

$$f_i(x) - f_{i+1}(x) = bx \text{ .} \tag{1.7}$$

This condition will prove to be helpful in the establishment of a computational procedure for the case involving concave functions.

For the general case in which the functions are concave and arbitrarily different for each period, setup costs also can be included in the expression (1.3) as follows:

Let the assumption be that if a plant or machine is put into operation during the i^{th} period, a fixed charge k_i is incurred. If it is not put into operation, the fixed charge is zero. Then, total costs are

$$q = \sum \left[f_i(x_i) + k_i\, \delta(x_i) \right], \tag{1.8}$$

where $\delta(x) = 1$ if $x > 0$

$\delta(x) = 0$ if $x = 0$.

Clearly, the discontinuity at the origin has no effect on the concavity of the function within the brackets. Therefore, (1.8) can be put in the form given by (1.3).

A. Concave Cost Functions

We consider first those problems in which the cost functions satisfy (1.1). The following definition will be convenient:

> A concave production plan is one in which a production schedule for "n" days consists of producing up to capacity for some "k" of the periods, producing a remaining fraction, "r", during one other period and for the remaining (n-k-1) periods producing nothing.

It is well known that a concave function defined over a convex set takes its minimum at an extreme point. Hence, if the cost functions are concave, we have

THEOREM 1: The basic scheduling problem is solved by the use of a concave production plan.

It seems to be a consequence of Theorem 1 that the development of an efficient computational procedure is a formidable problem. While the theorem reduces the search for a cost minimizing plan to a finite number of concave plans, it provides no further clues for the establishment of a suitable computational technique.

Special Cases: The existence of special cases, whose forms are such that they lead to simple computational schemes, provides one of the best means of extrication from the difficulties posed by the combinatorial aspects of the problem.

The most elementary of these cases occurs when the functions and the capacities are the same for each period. The total quantity to be scheduled can be written in the form, $S = kc + r$, where k is a non-negative integer and $r < c$. Then, an optimal plan is obtained by producing up to capacity for any k of the periods, producing the remainder, r, during any other period and producing nothing during the remaining periods. The proof follows directly from Theorem 1. The given plan is concave, and all concave plans are the same since all functions are the same.

It should be noted that no information concerning the parametric form of the cost function is required for planning purposes other than its concavity. Others [2,7] have noted a similar "minimal information" requirement in connection with planning problems involving a single convex cost function. It is shown later that this condition also holds for a slightly more general case concerning convex functions.

Another simple computational rule becomes available when the functions are such that the following conditions hold:

(2.9) $|f_i(x) - f_j(x)|$ is, for all i,j, a nondecreasing function of x as x increases

(2.10) $c_i = c$ for $i = 1, 2, \ldots, n$.

It may be recalled that the cost model, (1.6), developed for the basic scheduling problem fulfills condition (2.9); i.e., the condition expressed by (1.7) is a special case of (2.9). Condition (2.9) implies that the functions may be reindexed and ordered so that

(2.11) $$f_1 \leq f_2 \leq \ldots \leq f_n .$$

Assume that this has been done. Then the optimal plan is given by

Rule 1:

(a) Produce up to capacity for the periods associated with the first k cost functions where k is such that $kc < S \leq (k+1)c$.

(b) Produce the remainder, (r), during the period associated with the $k + 1^{st}$ cost function.

(c) Produce nothing during the remaining periods.

PROOF: Theorem 1 states that only concave production plans need be considered. It will be shown that any concave production plan involves costs which may be greater and which are never less than the costs associated with the plan obtained from the above rule.

It is clear from (2.9) that any other concave production plan that included the interchange of either a full capacity period or the $k + 1^{st}$ period in the proposed optimal plan with an idle period will involve higher costs. Hence, it is only necessary to consider those alternative plans which involve the utilization of the $k + 1^{st}$ period at full capacity and some earlier period at production level equal to r.

It is required to show that the following inequality holds:

$$f_i(c) + f_{k+1}(r) \leq f_{k+1}(c) + f_i(r) \tag{2.12}$$

where $i \leq k$

Upon rearrangement, (2.12) becomes

$$f_{k+1}(r) - f_i(r) \leq f_{k+1}(c) - f_i(c) . \tag{2.13}$$

Reference to (2.9) shows that (2.13) holds. Consequently, optimality of the plan given by the rule is demonstrated.

The rule does not always yield optimal plans if conditions (2.9) and (2.10) are not imposed. For many situations, however, its use with the modification described below may lead to "reasonably" good engineering results. If the cost of obtaining a plan by using a computer is contrasted with the ease of obtaining a "near" optimal plan, it may well be that an approximation of the sort proposed is sufficient.

The obvious modification required for approximation purposes is that (2.11) be replaced: (1) by an ordering related to costs at capacity, which is used to determine the k full-capacity periods and (2) by an ordering of the remaining functions. The latter is determined by the values of these functions at level r, which is used to select the $k + 1^{st}$ period. For the special case in which $S = kc$, where k is any positive integer, the modification clearly leads to an optimal solution.

B. Convex Cost Functions

It will be a convenience in this and subsequent sections to establish identifying nomenclature for a hierarchy of minimization problems which involve convex functions. The motivation for this step will become apparent.

The simplest problem of interest is:

Problem I: To minimize the function, q, of (1.3), subject only to the conditions (1.4) and

(3.1) f_i is a strictly convex[2] differentiable function.

The next problem was considered by Thrall [9].

Problem II: To minimize the function, q, of (1.3), subject to the restrictions (1.4), (3.1), and with negative values disallowed, i.e.,

$$x_i \geq 0 . \tag{3.2}$$

The third in the sequence is the basic scheduling problem.

Problem III: To minimize the function, q, of (1.3) with the subsidiary conditions (1.4), (3.1), and with an upper bound on each of the variables, i.e.,

$$0 \leq x_i \leq c_i . \tag{3.3}$$

A fourth problem will be described in a later section.

We first consider Problem I, to which a solution may contain negative components. The following well-known theorem is applicable. The proof follows immediately from the use of the method of the Lagrange multiplier technique.

THEOREM 2: A necessary and sufficient condition for X to be a cost-minimizing plan for the Problem I is that

$$f_i'(x_i) = \mu , \quad i = 1, 2, \ldots, n \tag{3.4}$$

where μ is a Lagrange multiplier.

Two equivalent methods are available to compute solutions. The first is analytical. Since f_i' is monotone, it has a unique inverse. If the inverse is denoted r_i, the conditions of the problem require that

$$S = \sum r_i(\mu) . \tag{3.5}$$

Equation (3.5) can be solved for μ. Given the value of μ, the value taken by each component of X can be computed by applying (3.4).

A graphical procedure also may be utilized:

If each function, f_i', is plotted, it is only necessary to find that horizontal line passing through all of the graphs of the functions such that the sum of the inverses of the functions at the points cut by the line sum up to the quantity S.

It will be shown that the above are also the only computational techniques required to solve Problems II and III.

---

[2] Strict convexity is defined for functions in which equality is obtained in the definition of convexity, (1.2), only when $x_1 = x_2$ for all values $0 < a < 1$.

We first state, without proof, the following result due to Thrall.

THEOREM 3 (Thrall): If each of the derivatives f_i' is continuous, then μ is a strictly monotone increasing function of S.

An immediate consequence of this is:

COROLLARY 1: If f_i is a strictly convex differentiable function and if X denotes an optimal solution to Problem I, then x_i $(i = 1, 2, \ldots, n)$ is a continuous strictly monotone increasing function of S.

The general computational procedure proposed will consist of solving first the unrestricted Problem I. If this solution is feasible (i.e., it meets all restrictions) for the higher numbered problems, clearly, no further work is required. If not, the following development will be utilized.

We first consider Problem II. Assume that an optimal solution to Problem I contains k negative components. The set of these negative components is denoted by K.

For our purposes, an important result is:

LEMMA 1: If X is optimal for Problem I, then for every arbitrary feasible plan Y there exists a feasible plan Z with $z_i = 0$, where $i \epsilon K$, such that

$$q(Z) < q(Y)$$

PROOF: We first show that for some value α, where $0 < \alpha < 1$ that Z can be constructed as a linear combination of X and Y.

Let

$$\alpha = \min_{i \epsilon K} \left[\frac{0 - y_i}{y_i - x_i} \right] \tag{3.6}$$

if

$$Z = \alpha X + (1 - \alpha) Y . \tag{3.7}$$

Then it is clear that $z_i = 0$, where i corresponds to that component for which α was found above.

Since a linear combination of feasible components has to yield a feasible component, we need only consider the effect on the remaining components of Z which are formed by using the nonfeasible (negative) components of X and the corresponding feasible ones of Y.

Let x_j denote a component for which $j \neq i$ and $j \epsilon K$. Then by (3.6)

$$\alpha \leq \frac{(0 - y_j)}{(x_j - y_j)} . \tag{3.8}$$

From (3.7) it is seen that

$$\alpha \leq \frac{0 - y_j}{\frac{z_j - y_j(1-\alpha)}{\alpha} - y_j}$$

which reduces to

$$z_j \geq 0 .$$

Hence Z is feasible.

The proof of the lemma is continued. By the strict convexity of q it is known that

$$q(Z) < \alpha\, q(X) + (1 - \alpha)\, q(Y) .$$

The desired result follows, provided that

$$q(X) < q(Y) .$$

This holds, since X represents a minimal cost plan.

The preceding result leads to

THEOREM 4: If Z is an optimal solution to Problem II, then $z_i = 0$ for each $i \in K$.

PROOF: The proof is outlined. Lemma 1 is first applied to show that for some $\hat{i} \in K$, $z_{\hat{i}} = 0$. Since an optimal solution must have this component at zero, the search for a cost-minimizing plan can be restricted to the subset of plans with $z_{\hat{i}} = 0$. This process can be repeated by finding the minimizing plan in this subset and applying Lemma 1 again to further reduce the size of the subset containing the optimal feasible plan. This is continued until a solution without negative components is obtained. The inclusion of all elements of K at the value zero in the final solution is shown by the following argument.

Let X denote the optimal solution to Problem I for scheduling the quantity S over periods 1, 2, . . . , n and X'' the optimal solution to Problem I for scheduling the quantity S over periods 1, 2, . . . , n - 1, where, for notational convenience, it is assumed that x_n was negative and that $\hat{i} = n$. Thus $x'_n = 0$. As such, it may be considered that the n^{th} period has been dropped from the problem.

Let X' denote the optimal solution to Problem I for scheduling the quantity $(S - x_n)$ over periods 1, 2, . . . , n - 1, i.e., X' is the plan obtained by eliminating the n^{th} component from X and leaving the other components unchanged.

Then, since $S < (S - x_n)$, by Corollary 1 it follows that $x''_i < x'_i$, for i = 1, 2, . . . , n - 1. Hence, if any components of X' are negative, their counterparts in X'' are also negative.

The foregoing is applicable to each stage of the construction outlined previously to obtain the optimal feasible solution. Consequently, all negative components of X eventually are set at the value zero.

The preceding work suggests the simple algorithm given below to solve Problem II.

Algorithm 1:

(a) Find the minimizing plan X for Problem I. If X is feasible for Problem II, the problem is solved. If not, proceed to (b).

(b) Set all negative components of X at zero value. Eliminate these periods from the problem and find the minimizing plan for the reduced problem by repeating step (a).

(c) Repeat steps (a) and (b) until step (a) yields a solution which is feasible for Problem II.

Termination of the algorithm is guaranteed, since each stage either yields an optimal solution or further reduces the dimensions of the next problem.

For Problem III the development follows essentially the same lines. If X now denotes an optimal solution to Problem II, which has a set K of components which exceed their capacities, then our stepping stone is:

LEMMA 2: If X is not feasible for Problem III, then for every arbitrary feasible plan Y there exists a feasible Plan Z, with $z_i = c_i$, where $i \epsilon K$, such that

$$q(Z) < q(Y) .$$

PROOF: The proof follows the same lines as that of Lemma 1. It is noted first that Z can be constructed as

$$Z = \alpha X + (1 - \alpha) Y .$$

Z is feasible if α is chosen as follows:

$$\alpha = \min_{i \epsilon K} \left[\frac{c_i - y_i}{y_i - x_i} \right].$$

The remainder of the proof is the same, except for obvious alternations, as that given in Lemma 1, and will not be repeated.

The above generalizes to:

THEOREM 5: If Z is an optimal solution to Problem III, then $z_i = c_i$ for each $i \epsilon K$.

Proof of this theorem follows much the same arguments as Theorem 4, and therefore is omitted. It forms the basis for the computational procedure given below for Problem III.

Algorithm 2:

(a) Compute an optimal solution for Problem II. If this plan is feasible for Problem III, it is optimal, and the problem is solved; if not, proceed to (b).

(b) Reduce the value of each component in the above plan which exceeds its capacity restriction to the value of its capacity. Then, eliminate these components from the problem and repeat step (a) for the new problem with

$$S' = S - \sum_{i=1}^{k} c_i ,$$

where c_i $(i = 1, \ldots, k)$ represents the capacity of the components eliminated from the original problem.

(c) Repeat (a) and (b) until a feasible plan is obtained.

As noted previously, both Modigliani and Hohn and Charnes, Cooper, and Mellon considered the case in which inventory costs were negligible, and the convex cost function was the same for all periods. The decision rule for these conditions takes the simplest possible form: to divide the total quantity to be produced evenly among all of the periods. Further, it was shown that this rule applies for any convex cost function, quite independently of its particular parametric form.

This same rule can be obtained for the slightly more general condition given in:

THEOREM 6: A minimal cost production plan for the problem defined by (1.2), (1.4), (1.5), and

$$(3.9)^3 \qquad f_i(x) - f_j(x) = d_{ij} \quad \text{for all } i, j = 1, 2, \ldots, n$$

$$\text{and for all } 0 \le x \le \frac{S}{n} \le c_i$$

is given by:

$$\hat{x} = x_i = \frac{S}{n} \qquad \text{for } i = 1, 2, \ldots, n .$$

PROOF: Let x, $i = 1, \ldots, n$ be arbitrary. Then, by hypothesis

$$\hat{x} = \frac{S}{n} = \frac{1}{n} \sum_i x_i ,$$

by the convexity of f_j

$$f_j(\hat{x}) \le \frac{1}{n} \sum_i f_j(x_i) .$$

Summing over j, yields

$$\sum_j f_i(\hat{x}) \le \frac{1}{n} \sum_j \sum_i f_j(x_i) .$$

---

[3] This condition may occur, for example, in situations in which the learning characteristics of a labor unit is known to follow some known "learning curve." Examples of this sort are known to exist in the aircraft industry [8].

Rearrangement of the right hand side and the application of (3.9) results in

$$\sum_j f_j(\hat{x}) \leq \sum_j f_j(x_j) ,$$

completing the proof.

For the case in which some $c_i < \hat{x}$, Algorithm 2 may be applied.

IV. SCHEDULING A SERIES OF REQUIREMENTS

Supplementing the notation utilized for the basic scheduling problem, let

S_i denote the minimal product quantity required to be available by the end of the i^{th} period, where $i = 1, 2, \ldots, n - 1$, and

S_n denote the total quantity required to be available by the end of the n^{th} period.

Under the assumption of linearity for inventory charges, total costs are given by

$$q = \sum_{i=1}^{n} h_i(x_i) + \sum_{i=1}^{n-1} b \left(\sum_{j=1}^{i} x_j - S_i \right)$$

where: $h_i(x_i)$ denotes the cost of production during the i^{th} period,

b the unit inventory cost per unit of time.

The latter term in the above expression arises under the assumption that inventory charges exist for production in excess of requirements only. That is, it is assumed that items are shipped or used according to the requirement schedule.

The above may be reduced to

$$q = \sum_{i=1}^{n} \left[h_i(x_i) + b \left((n - i)x_i - S_i \right) \right] .$$

If

$$f_i(x_i) = h_i(x_i) + b \left((n - i)x_i - S_i \right) ,$$

the total cost function takes the form (1.3).

Since f_i is the sum of a concave (convex) function and a linear function, it is concave (convex).

It may be noted that the constraint of exactly meeting and not exceeding the cumulative requirements for the last period has been retained in this extension of the basic scheduling problem. This was done in order to keep the formulation relative to the cost functions as general as possible. If the cost functions are restricted to be monotonically increasing, or at least nondecreasing, this terminal restriction can be weakened to an inequality. This would

not vitiate the results in any way since the same computational procedures are applicable to either problem. Production costs, as a function of output, are well known to be monotonically increasing functions, hence, the scheduling problem can be formulated either way.

A. Convex Cost Functions

Problem IV: To minimize the function, q, of (1.3) subject to the restrictions (1.4), (3.1), (3.3), and the additional qualifications given below, related to the necessity of meeting a series of requirements.

$$\sum_{j=1}^{i} x_j \geq S_i \qquad \text{for } i = 1, 2, \ldots, n-1 \tag{4.1}$$

where

$$S_i \leq \sum_{j=1}^{i} c_j .$$

The horizon planning method given by Modigliani and Hohn is suggestive of a computational technique for this problem.

Algorithm 3:

(a) Solve Problem III for scheduling the quantity S_n over periods $1, 2, \ldots, n$. If this plan is feasible for Problem IV the problem is solved. If not, proceed to step (b)

(b) Compute an optimal feasible solution to Problem III for scheduling the quantity S_{k_1} over the periods $1, 2, \ldots, k_1$, where k_1 is the latest period for which such a solution can be found.

(c) Reduce the value of each requirement after the k_1^{st} period by the quantity S_{k_1} and repeat step (a) for the new problem of scheduling $(S_n - S_{k_1})$ over periods $k_1 + 1, k_1 + 2, \ldots, n$.

(d) Continue steps (b) and (c) until a plan which is feasible for Problem IV is obtained. This plan is also the minimal cost solution.

PROOF: Assume that a plan (Z) has been obtained via the proposed algorithm. It will be shown that any perturbation of this plan leads to one involving higher costs.

Z can be considered to have the following appearance:

$$\left[Z_1, \ldots, Z_{k_1}\right], \left[Z_{k_1+1}, \ldots, Z_{k_2}\right], \ldots, \left[Z_{k_{r-1}+1}, \ldots, Z_n\right]$$

where the brackets are used to separate the r independent sub-problems formed by the algorithm.

Let μ_j $(j = 1, 2, \ldots, r)$ denote the common value of the partial derivatives of the positive and less than capacity components within each sub-problem.

It is necessary to digress and first show that the following relation exists:

$$\mu_1 > \mu_2 > \ldots > \mu_r . \tag{4.2}$$

It may be recalled that by Theorem 3 the constant μ is a monotonic increasing function of S.

Consider the effect of the adjustment given by steps (b) and (c) of the algorithm. If $\sum_{i=1}^{k_1} x_i$ represents the cumulative sum of the components of the original solution, X, obtained in step (a), then, since the reallocation required to meet the restriction S_{k_1} was an increase, i.e., $S_{k_1} > \sum_{i=1}^{k_1} x_i$, it follows that

$$\mu_1 > \mu ,$$

where μ is the constant associated with the plan, X.

On the other hand, if μ_1' is the constant associated with the solution to scheduling $(S_n - S_{k_1})$, then, since $(S_n - S_{k_1}) < \sum_{k_1+1}^{n} x_i$, it follows that

$$\mu_1' < \mu$$

and therefore

$$\mu_1 > \mu_1' .$$

If the solution to scheduling $(S_n - S_{k_1})$ is feasible, μ_1' may be denoted μ_2' and the asserted relationship holds. If the solution is not feasible, the above argument is repeated for the sub-problems formed by applying steps (b) and (c) once again. The constants associated with these sub-problems are related to μ_1' as before.

That is, $\mu_2 > \mu_1' > \mu_2'$.

If the solution to the latter of the sub-problems is feasible

$$\mu_2' = \mu_3$$

and

$$\mu_2 > \mu_3 , \mu_1 > \mu_3 .$$

It remains to show that $\mu_1 > \mu_2$. Consider the problem of scheduling the quantity S_{k_2} over periods 1, 2, ... k_1, ..., k_2. Let $\bar{\mu}$ denote the value associated with the solution to the problem. Since this solution is (by hypothesis) not feasible with respect to the requirement of period k_1, an adjustment is required. The nature of the adjustment is to increase the quantity scheduled over the periods up to k_1 and to decrease the quantity scheduled over the remaining periods. By previous arguments this leads to the relationship, $\mu_1 > \bar{\mu} > \mu_2$. Thus, the ordered sequence given by (4.2) holds for the first three terms. The remainder of the relationship

given by the sequence may be obtained by similar arguments if there are more than three sub-problems in the final solution.

To continue with the proof, it will now be shown that any perturbation of the given plan in a permissible direction leads to higher costs. First, it is clear that any changes within each sub-solution are of no advantage since each sub-solution was optimal for the periods and quantity scheduled. The only other permissible change is to increase the quantity to be scheduled in one interval and to decrease it in another as compensation. In order to maintain feasibility, the increase must be in the earlier of the two periods.

In order for the plan obtained via the algorithm to be optimal, it is required to show that

$$\sum_a f_i(z_i) + \sum_b f_i(z_i) < \sum_a f_i(z_i + e_i) + \sum_b f_i(z_i - e_i) \tag{4.3}$$

where a, b denote arbitrary different sub-problems formed by the algorithm, a < b

and $$\sum_a e_i = \sum_b e_i \ , \ (e_i > 0) \ .$$

Consider the two expressions:

$$\sum_b \left[\frac{f_i(z_i) - f_i(z_i - e_i)}{e_i} \right] \frac{e_i}{\sum_b e_i}$$

and

$$\sum_a \left[\frac{f_i(z_i + e_i) - f_i(z_i)}{e_i} \right] \frac{e_i}{\sum_a e_i} \ .$$

Letting e_i's go to zero, the first expression becomes μ_b and the second, μ_a.

It was previously shown that $\mu_b < \mu_a$. Hence, for e_i's sufficiently small, it follows that

$$\sum_b \left[\frac{f_i(z_i) - f_i(z_i - e_i)}{e_i} \right] \frac{e_i}{\sum_b e_i} < \sum_a \left[\frac{f_i(z_i + e_i) - f_i(z_i)}{e_i} \right] \frac{e_i}{\sum_a e_i}$$

which implies (4.3).

B. Concave Cost Functions

Following the earlier discussion related to the combinatorial aspects of the basic scheduling problem for the case of concave functions, the work in this section also will be devoted to computational aspects of two special cases of interest.

As in the basic scheduling problem, the most elementary of these occurs when the same cost function exists for every period. This presents no computational difficulties. The solution, for obvious reasons, is merely to set up any concave plan which fulfills all requirements.

The second is also the same as that considered in the section on the basic scheduling problem for concave functions. The only addition being the additional restrictions imposed by the condition (4.1) of meeting or exceeding an intermediary series of requirements.

A complete description of this problem is to minimize the function q of (1.3) subject to conditions (1.1), (1.4), (1.5), (2.9), (2.10), and (4.1). For convenience, this will be designated as Problem IIa, while the basic scheduling problem, omitting condition (4.1), will be referred to as Problem Ia. The computational procedure suggested is given below.

Algorithm 4:

(a) Obtain an optimal solution for Problem Ia using Rule 1. If this plan is feasible for Problem IIa, it is also an optimal plan. If it is not feasible, proceed to step (b).

(b) Construct a new plan as follows:

If S_{k_1} is the earliest requirement which is not met by the above plan, obtain by using Rule 1 an optimal solution for the problem of scheduling S_{k_1} over periods $1, 2, \ldots, k_1$. Then, compute solution using Rule 1 for the problem of scheduling $\left(S_n - S_{k_1}\right)$ over periods $k_1 + 1, k_1 + 2, \ldots, n$, with the requirements suitably adjusted by subtracting S_{k_1} from each requirement within this interval. The first plan so obtained must be feasible. If the second plan is also feasible, then the combined plan is optimal. If the latter plan is not feasible, proceed to step (c).

(c) Obtain using Rule 1:

(1) The solution for scheduling the quantity, S_{k_2}, over periods $1, 2, \ldots, k_2$, where S_{k_2} is obtained similarly to S_{k_1}, subject to the restriction that no components of the optimal plan for scheduling S_{k_1} may be decreased.

(2) The solution to scheduling the quantity, $(S_n - S_{k_2})$, over the remaining periods with the requirements suitably adjusted by subtracting S_{k_2} from each requirement within this interval.

(d) Continue step (c) until both plans thus obtained are feasible.

A proof is somewhat long-winded because of the combinatorial aspects of the problem. However, its spirit is easily summarized. If a feasible plan is not obtained from step (a) of the algorithm, then some adjustment must be made in the way of reallocating production to some higher-cost earlier periods from some lower-cost later periods. Since, the direction of the reallocation is unfavorable, the desire is to reallot the least possible quantity. The restriction in step (c2) is required to maintain feasibility.

A production scheduling rule which may be stated in a most simple form exists for the following problem.

If (a) condition (1.7) holds, and

(b) the cumulative requirement schedule is such that

$$(S_k - S_j) \leq (k - j)\, c \quad \text{where} \quad k > j\,,$$

then the rule is to always set back production sufficiently to just meet and never to exceed the requirements.

The proof follows immediately from the application of Algorithm 4 and the timewise inverse ordering of the cost functions caused by (1.7).

V. CONCLUDING REMARKS

In addition to the direct application of the methods of solution to the production-planning problems originally posed, translations of these solutions can, for some problems, be made in the form of simple operating guides. These guides may be of use at different management levels to help cope with the exigencies of routine operations. For the basic scheduling problem, assuming that the total interval is short so that inventory charges and other known cost changes may be ignored, and if the nature of the cost functions is known, the guide—in terms of some simple relationship among the periods within an interval—could be expressed as a function of the total quantity to be scheduled. If the daily production cost function is convex for a segment of a plant, the manager of such a section knows how to reschedule daily production in the face of changes from previously specified requirements. If the change is a decrease, the rule is to decrease each daily schedule by an even quantity. Similarly, for an increase, the allocation also is equal. It may be noted that Theorem 6 implies the validity of this rule for other cases in which the convex cost functions differ by a constant.

On the other hand, if the daily cost function was concave, the guide would be to maintain full capacity production for as many days of the week as it required.

As previously noted, the solution to the problem of scheduling to meet a series of requirements was shown by Modigliani and Hohn to give rise to an interesting phenomenon. They demonstrated for the case in which the production cost function was convex and the same for all periods, as well as for linear inventory costs, that production had only to be planned for a certain number of periods in advance. The optimal production schedule was found to decompose into a series of independent sub-schedules. It was shown that this same phenomenon occurs in the more general case involving arbitrarily different, strictly convex, cost function. This provides further support for established management practices of planning for the "immediate future." But it is clear that for optimal planning the extent of the immediate future is not arbitrary.

Of equal interest is the solution to the elementary production-scheduling problem involving concave cost functions which are related according to (1.7). This is the analogous problem involving a concave rather than a convex production cost function which is invariant in time and linear inventory charges to that considered by Modigliani and Hohn. The solution to this problem given by Algorithm 4 indicated that production has to be planned only to just meet each requirement. Thus, the optimal production plan not only decomposes into a series of independent sub-plans, but the points at which these decompositions take place are a function only of the requirement schedule. This solution is also independent of the exact form of the production cost function.

In many industrial activities the difficulty of determining the analytical form of such cost functions as have been under consideration is well known. It is clear, however, that if decisions are to be made in a rational fashion, with cost minimization or profit maximization as major choice principles, such information in the form required by decision-making procedures must be obtained. It has been pointed out that for some production planning problems, a modicum of such information is required.

ACKNOWLEDGMENTS

I am indebted to Professor C. Derman for directing my attention to this problem area. Without his clarifying suggestions and penetrating criticism this study could never have been completed. Professor S. B. Littauer also provided many encouraging and fruitful discussions for which I am deeply grateful.

REFERENCES

[1] Antosiewicz, H., and A. J. Hoffman, "A Remark on the Smoothing Problem," Management Science 1 (1954)

[2] Charnes, A., W. W. Cooper, and B. Mellon, "A Model for Optimizing Production by Reference to Cost Surrogates," Econometrica 23 (1955)

[3] Hoffman, A. J., and W. Jacobs, "Smooth Patterns of Production," Management Science 1 (1954)

[4] Holt, C. C., F. Modigliani, and J. F. Muth, "Derivation of a Linear Decision Rule for Production and Employment," Management Science 2 (1955)

[5] Holt, C. C., F. Modigliani, and H. A. Simon, "A Linear Decision Rule for Production and Employment Scheduling," Management Science 2 (1955)

[6] Johnson, S. and G. Dantzig, "A Production Smoothing Problem," Proceedings of the Second Symposium in Linear Programming, Washington, D. C., January 1955

[7] Modigliani, F., and F. E. Hohn, "Production Planning over Time and the Nature of the Expectation and Planning Horizon," Econometrica 23 (1955)

[8] Raborg, W. A., Jr., "Mechanics of the Learning Curve," Aero Digest 65 (1952)

[9] Thrall, R. M., "Some Results in Non-Linear Programming," Proceedings of the Second Symposium in Linear Programming, Washington, D. C., January 1955

* * *

SIMULATION RESEARCH ON JOB SHOP PRODUCTION[1]

James R. Jackson*
University of California, Los Angeles

1. INTRODUCTION

This paper reports simulation research on certain problems of industrial production. The computations were carried out by W. H. Marlow, on the Logistics Computer, Logistics Research Project, George Washington University.

The particular situation studied is a miniaturized idealization of the machine shop of a plant which manufactures small lots of precision units to customer order and specifications. Our interest is focussed on systematic procedures for answering the question "What should this machine do next?" The general objective is assumed to be on-time completion of orders, relative to given due-times. The problem studied is set up in such a way that over-all on-time completion is expected to be impossible, and the specific objectives considered are concerned with minimizing certain functions of tardinesses.

Section 2 describes the production situation studied and outlines the computations by presenting a paper-and-pencil model. Section 3 summarizes the numerical inputs for the computations. Section 4 outlines preliminary manual experiments and discusses the procedures studied for deciding what a machine should do next. Section 5 presents and discusses computational results.

2. THE MACHINE SHOP MODEL

The reader should have no difficulty in relating the undefined terms in the following description of our model to their referents in the machine shop itself. The physical constituents of the model are as follows.

A. Several two-compartment boxes, the compartments being "waiting lines" and "machines."

B. For each machine, a half line called its "time axis," marked in discrete units starting with zero.

C. A box called the "transport system" and a positive integer T called the "minimum move-time."

---

*Manuscript received February 19, 1957

[1]Prepared under the title, "Job Shop Simulation by the Logistics Computer," as part of the Management Sciences Research Project, University of California, Los Angeles, under Office of Naval Research Contract Nonr-233(02). Many persons have contributed to this research, especially R. T. Nelson, A. J. Rowe, and W. H. Marlow. The author, of course, bears the entire responsibility for errors and shortcomings in the present report.

D. Cards called "jobs." On each job is listed a "due time," and a sequence of machines. Opposite each machine is a positive-integer-valued distribution, called the "processing-time distribution" for the "machining operation" represented by the listing of machine on the job. The jobs also have a blank line for recording "ready times."

E. Scratch paper, pencils, etc., a mechanism for choosing random values from the above-mentioned distributions, and a "scheduler."

Cards are stored in the boxes and moved around to simulate the movement of production lots through a shop. The model is operated in a synthetic time whose actual rate of advance is controlled by the scheduler.

Figure 1 shows a flow chart for the computations. The output is a bar chart showing the history of the simulation. This permits the comparison of simulated histories on the basis of a wide variety of criteria.

We briefly outline the significance of the main blocks of the flow chart. At the start, the jobs are set up in waiting lines appropriate for their first machining operation (Block A, Figure 2). Then jobs are assigned to machines, and the "actual" machine times to be needed are blocked off on the time axes (Block D). Then the synthetic time is advanced (Block E), and selected jobs are shifted from the transport system to appropriate waiting lines (Block B). Then all jobs whose operations have been completed are removed from machines, and go to discard if "finished," otherwise to the transport system on the way to the next required machine (Block C). At this point the cycle starts over again with Block D and continues until all jobs are completed.

It is necessary to insert a criterion in Block D for deciding which of several available jobs is to be selected. Each of the criteria to be studied here is based on a specified function of job data, the rule being to choose the job for which the functional value is least. The particular functions considered will be discussed in Section 4.

A comment is needed to explain the treatment of machining operations as indivisible entities. The jobs are intended to represent quantity lots, so the performance of a given machining operation could actually be split into two or more time intervals in the shop itself. But the need for setting up special attachments and making adjustments places a penalty upon such splits. In some shops this penalty is large enough so that splits are rarely feasible; the reader with an urge for realism may think of the present model as representing such a shop. From another point of view, though, such restrictions on alternatives as the elimination of splits can be interpreted as indications of what management choices are to be considered as exceptions to standard procedures, rather than as blanket limitations on the shops to which findings concerning the model may be applicable.

Another gross simplification lies in the crude treatment of transport, which we ignore, except for the specification of T. Generalizations in this regard are planned for future research.

Additional special characteristics of the model, most of which will be generalized in fugure work, are the designation of a single specific machine for each machining operation; the availability of all jobs at time zero; the avoidance of complications resulting from spoilage and engineering changes; the implicit requirement that all work be done by the shop itself and in regular work hours; and the implicit assumption that personnel, tooling, and raw material needs will not effectively limit the shop's behavior.

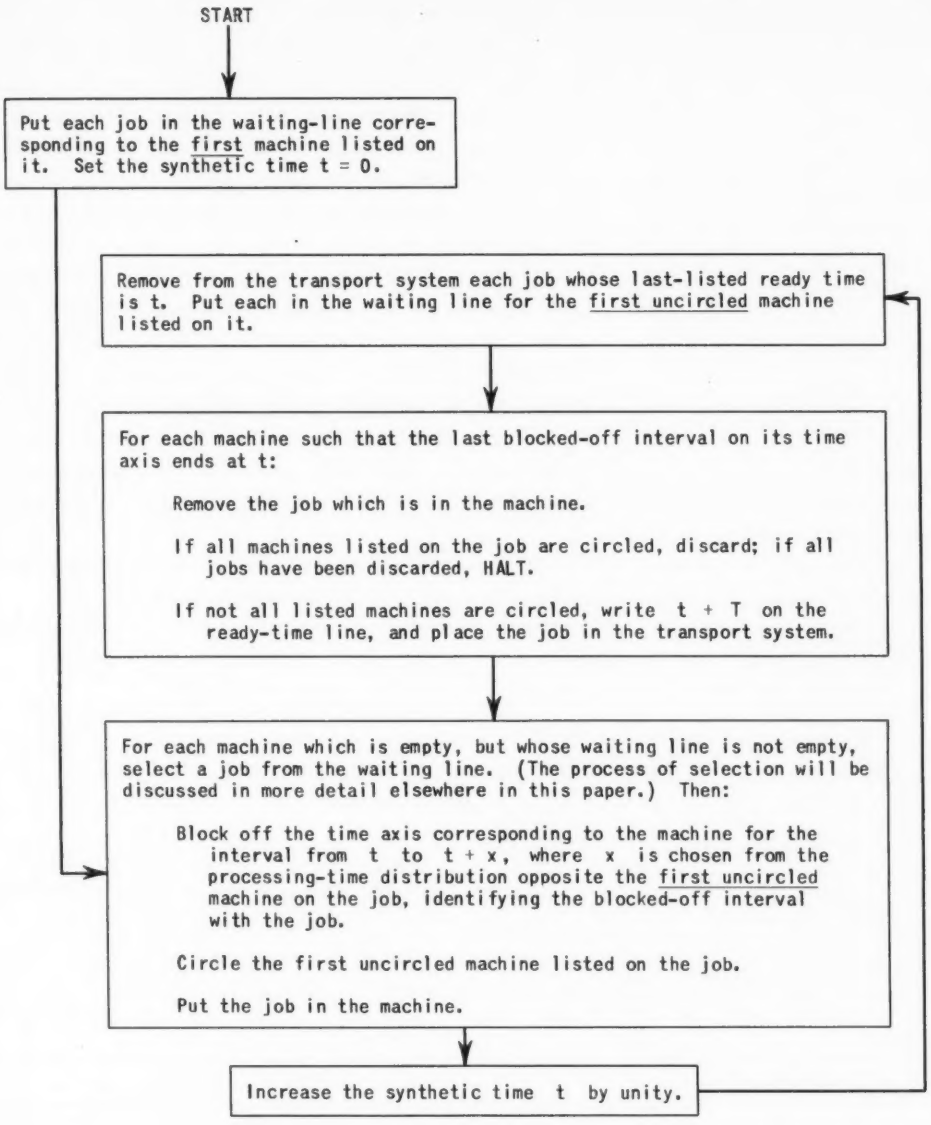

START
Put each job in the waiting-line corresponding to the first machine listed on it. Set the synthetic time t = 0.
Remove from the transport system each job whose last-listed ready time is t. Put each in the waiting line for the first uncircled machine listed on it.
For each machine such that the last blocked-off interval on its time axis ends at t:
Remove the job which is in the machine.
If all machines listed on the job are circled, discard; if all jobs have been discarded, HALT.
If not all listed machines are circled, write t + T on the ready-time line, and place the job in the transport system.
For each machine which is empty, but whose waiting line is not empty, select a job from the waiting line. (The process of selection will be discussed in more detail elsewhere in this paper.) Then:
Block off the time axis corresponding to the machine for the interval from t to t + x, where x is chosen from the processing-time distribution opposite the first uncircled machine on the job, identifying the blocked-off interval with the job.
Circle the first uncircled machine listed on the job.
Put the job in the machine.
Increase the synthetic time t by unity.

Figure 1

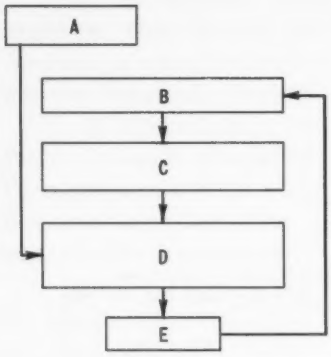


Figure 2

3. NUMERICAL INPUTS

The "shop" studied has 8 machines and was to process 70 jobs, requiring a total of 263 machining operations, each job needing 2 to 5 operations. The numerical inputs to be described below were chosen in the attempt to imitate, in miniature form, typical situations prevailing in certain aircraft-parts machine shops with which this writer has had contact.

The number of machining operations required by a given job, the actual sequence of machines, and the average values for the processing-time distributions were chosen by random processes. These processes were designed to result in three equally loaded "bottleneck" machines, whose loads were considerably more than those of the other machines; to result in some machines performing many short operations, and others performing fewer but longer operations; and to produce a slight tendency for jobs to "flow" from low-numbered to high-numbered machines.

The mean of the average values of the processing-time distributions was 9.90. These average values were all multiple of 3, ranging from 3 to 21. Nine choices of "actual" processing times were made as follows (in each choice, the selection of "actual" times was made independently for each operation):

Choice 1. Actual = average value = m

Choices 2, 3, 4, 5. Actual =

2m/3	with probability	0.3
m	" "	0.5
4m/3	" "	0.1
5m/3	" "	0.1

Choices 6, 7, 8, 9. Actual =

2m/3	" "	0.6
m	" "	0.1
4m/3	" "	0.1
5m/3	" "	0.1
2m	" "	0.1

Total "actual" times for the machining operations assigned to each machine according to each of these nine choices are given in Figure 3. Minimum move-time T = 3 was used throughout.

The process for choosing due times was slightly complicated, because in order to have a meaningful problem it is necessary that due times be within reason but not easy to achieve. We actually designed the problem so that it was expected that a substantial number of past-due completions would be unavoidable. The due times were chosen by ordering the jobs at random, then calculating for each job the time at which it would be completed if all the machining operations for it and for preceding jobs were done in their average times by continuous application of all machines, then increasing the smallest times so obtained by about 50 and larger times by systematically smaller and smaller amounts, and finally perturbing the resulting times randomly.

Figure 4 gives the basic data for a sample from the 70 jobs.

Number of Machine	Choices 1	2	3	4	5	6	7	8	9
1	240	232	239	229	228	274	219	253	231
2	303	303	318	288	324	299	331	320	249
3	396	383	442	390	428	330	377	389	350
4	285	280	264	265	295	267	263	271	306
5	396	374	429	367	407	433	365	408	408
6	297	309	309	277	285	367	292	324	338
7	396	377	385	354	406	431	384	398	377
8	291	295	304	306	324	291	313	262	255
Totals	2604	2553	2690	2476	2697	2692	2544	2625	2514

Figure 3 - "Actual" time gross statistics. Underlines indicate most heavily loaded machines ("bottlenecks").

Job No.	1	20	54	62
Due Time	108	276	243	366
Machine Sequence	1 6 8	3 2 7 8	6 1 7 8 4	7 1
Average times	6 18 18	6 15 12 18	15 9 6 6 9	6 9

Figure 4 - Sample jobs

4. PREVIOUS EXPERIMENTS; METHODS OF CHOICE

Prior to the experimentation reported here, a substantial amount of hand computation on a wide variety of extremely small problems had been carried out, by R. T. Nelson and by this writer. This work was mainly concerned with determining what sort of decision criteria would be useful for deciding what job to assign next to a given machine. This work was of an exploratory nature, and there would be no point in writing a full report on it.[2] Suffice it to say that, for all the problems studied, decision criteria of the type about to be presented provided the best results (with the exception of some iterative schemes, which have been rejected because of their head-on conflict with certain practical aspects of scheduling). Incidentally, these criteria usually yielded better results than experienced schedulers could get, using judgment to make decisions, but this may be due partly to the artificiality of the problems studied.

The criteria which turned out best in the preliminary computations, and which are used for the selection process of Block D of the present computations, are of the following form: "Choose the job from among those available for which p_{ab} is least," where

[2] Another brief summary is included in A. J. Rowe and J. R. Jackson, Research Problems in Production Routing and Scheduling, Journal of Industrial Engineering, May-June 1956.

$$p_{ab} = \text{(due time)} - \text{(time required for future moves)} - a \times \text{(expected time required for future machining)} + b \times \text{(expected time for present machining operation)}$$

In the present experiments, the p_{ab} were calculated for all combinations of a = 1, 2, 3, and b = 0, 1; and also for a = 0, b = 0. Each of these seven decision criteria was run against each of the nine choices of "actual" times.

A word is in order concerning the logic behind the p_{ab}. The number $p_{1,-1} - t$ for a job is simply the amount of time it can wait (neither in a machine nor in transport) and still be finished on time, if machining operations require their average times. This is obviously a "reasonable" criterion of choice, and in certain types of very small deterministic problems it can be shown to be mathematically correct. Dropping t has no effect upon the results of comparison, since t is the same for all jobs when a comparison is made. Moderate increases in the parameter a can be justified intuitively as taking into account the expectation of enforced waits occurring when a job arrives at an already occupied machine and also as taking into account the desirability of performing operations which will release a lot of work to other machines in the shop. Increases in b are harder to justify, although b = 0 is mathematically correct for some types of small problems. The use of b = 0, 1 in the present experiments is based upon the success of these values in previous experimental computations.[3,4]

5. RESULTS AND CONCLUSIONS

Since our problem is designed so we do not expect over-all on-time completion (and hence cannot simply differentiate between "good" and "bad" results), we must choose objective functions as bases for comparing the histories resulting from various simulation runs. The "completion time" of a job is the upper end of the last interval blocked off for it on a machine axis (which, of course, corresponds to the last machine required by the job). The job's "tardiness" is its

(Completion-time) - (Due-time),

if this difference is positive; otherwise the job's tardiness is zero. We considered the following three objective functions:

L_1 = (Sum of tardinesses of the jobs)

L_2 = (Sum of squares of tardinesses of jobs)

$M = \sqrt{L_2}$

Except for minor quantitative details, the results were the same for all three objective functions.

The basic problem was to determine the importance of changes in a and b, and what (a, b) combinations give the best results.

---

[3] Some of the complications arising in theoretical studies of job-lot scheduling are indicated by this writer's paper, "Scheduling a Production Line to Minimize Maximum Tardiness," Management Sciences Research Project Research Report No. 43, UCLA, 1955 (mimeographed).

[4] Experimentation now under way suggests that the choices b = 0, 1 may have been premature; and that the combination, a = 1, b = -1, is near best under fairly broad circumstances.

Figures 5 and 6 give the values for the various runs of $L_1 - \overline{L_1}$ and $L_2 - \overline{L_2}$, where $\overline{L_1}$ and $\overline{L_2}$ are the grand means of L_1 and L_2 over all runs. A similar table for M is omitted, since it shows nothing essentially new, and can be obtained from Figure 6 if desired.

a	b	Choices 1	2	3	4	5	6	7	8	9	Row average
0	0	-64	-176	+721	-267	+475	+41	-86	+122	+44	+90
1	0	-193	-306	+709	-325	+343	-96	-237	+93	-146	-18
1	1	-126	-306	+548	-369	+156	-104	-264	-200	-125	-88
2	0	-172	-346	+744	-317	+327	+10	-244	-188	-118	-34
2	1	-175	-354	+657	-310	+378	-67	-230	-127	-141	-41
3	0	-67	-287	+819	-190	+585	+199	-225	-16	-31	+87
3	1	-139	-301	+796	-286	+376	+19	-230	-58	-163	+2
Column average		-134	-297	+713	-296	+377	0	-217	-53	-97	

Figure 5 - $L_1 - \overline{L_1} = L_1 - 478$

a	b	Choices 1	2	3	4	5	6	7	8	9	Row average
0	0	-54	-89	+496	-127	+312	+39	-98	+86	+30	+66
1	0	-124	-152	+421	-149	+208	-87	-142	-74	-101	-22
1	1	-133	-144	+293	-164	+113	-98	-147	-109	-101	-54
2	0	-115	-157	+504	-155	+189	-57	-141	-116	-100	-16
2	1	-115	-160	+437	-153	+202	-67	-145	-94	-89	-20
3	0	-81	-146	+574	-104	+379	-11	-134	-55	-73	+39
3	1	-94	-153	+600	-133	+226	-43	-133	-80	-101	+10
Column average		-102	-143	+475	-141	+233	-46	-134	-63	-76	

Figure 6 - $10^{-2} \times (L_2 - \overline{L_2}) = 10^{-2} \times L_2 - 186$

A two-factor analysis of variance[5] showed that the row differences in Figure 5 are highly significant, with an F-ratio of about 10, corresponding to a significance level greater than 99.9 percent. Thus, we can conclude that there is a real difference among the results given by the various (a, b) combinations. Parallel analysis for the data of Figure 6 and for the table of M values give essentially the same results.

When we look at the tables, however, it is obvious that the column differences are much more important than the row differences. More specifically, in Figure 5, the standard deviation

[5]Mood, A., Introduction to the Theory of Statistics, McGraw-Hill, 1950, pp. 329-334.

of the column means from the grand mean is about 300, while the standard deviation of the row means is only 20. This preponderance of the column differences is also indicated by the fact that the formal application of a one-factor analysis of variance (treating the values of L_1 corresponding to a given choice of a and b simply as a sample of nine) shows no significant effect of rows. The exact interpretation of this result seems to be that the effect of variation in a and b is so strongly dominated by the effect of chance occurrences that it could not be detected without the unusually strong sort of replication that simulation allows—except at the expense of a very large number of observations. Similar analysis has not been carried out in full for the tables of L_2 and M, but it is clear that the results would be qualitatively the same.

In spite of the fact that the row differences might in practice be masked by chance occurrences, they could still be of economic importance. Hence we carry the analysis of these differences further.

First, when we look at Figures 5 and 6, it is clear that the row with a = 1, b = 1 is best with respect to either criterion. More generally, though, there seems to be little choice among the four rows with a = 1, 2. But the difference between the averages of the two of these with b = 0 and the two with b = 1 is highly significant statistically (again on the basis of a two-factor analysis of variance), and it is clear that the difference is in favor of the rows with b = 1.

We do not wish to over-generalize these conclusions but to point out that they can be given wider credence than the present experiment would allow by itself, since they verify informal hypotheses based upon the preliminary hand computations concerned with a wide variety of problems.

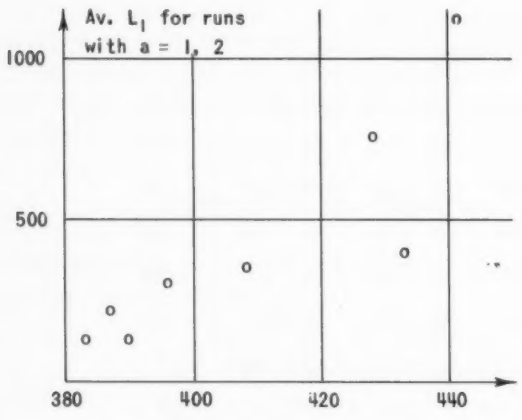


Figure 7

We also looked into the problem of predicting the column differences on the basis of gross input statistics. A large number of input statistics were examined, and, while statistical tests could not be soundly applied, it was obvious that the best predictor among those examined was the maximum of the total "actual" times required by any one machine to perform all the machining operations required of it. Figure 7 shows the average L_1 over the four rows with a = 1, 2 plotted against this indicator. Of course little confidence can be placed in this result, which is based upon so little data, and further experiments will be planned toward testing it.

In addition to comparing the results of using various values of a and b, it was desired to determine whether there would be any variation in the outputs attributable to the difference in the ratio σ/μ, of standard deviation to mean, for the two distributions used to obtain "actual" times. This ratio is 0.30 for choices 2 - 5 and 0.47 for choices 6 - 9; and it was felt that the more accurate detailed predictions going into the p_{ab} in the first case might lead to better outputs. But the data provide no substantiation for this idea.

6. VALIDATION: AN OPEN QUESTION

The reader who is familiar with the enormous complexity of real-life machine shops will recognize that the word "simulation" in the title of this article is used almost metaphorically. The computations did not imitate the operations of any specific shop, and are in fact concerned with a system of a different order of magnitude than is typical of the metalworking industry. Of course analogous systems of small size can be found in other areas. But if we restrict ourselves, for the present at least, to potential machine-shop applications, an all-important question stands out: To what extent can the results of computations such as are reported here be transferred to real machine shops?

The answer to the above question can be clinched only by trial—and possibly error, and potentially expensive error, too. A preliminary question, however, is this: How and in what sense can the type of model considered here (miniaturized, as well as idealized) be validated, short of direct application? Any suggestions will be much appreciated.

* * *

MATHEMATICAL PROGRAMMING AND EMPLOYMENT SCHEDULING*

W. Karush
The University of Chicago and
The Ramo-Wooldridge Corporation

and

A. Vazsonyi
The Ramo-Wooldridge Corporation

1. INTRODUCTION

We are concerned with the problem of planning employment levels w_i (number of workers) of a given labor type over an interval in time of N periods $i = 1, 2, \ldots, N$. It is supposed that the productive manhour requirements r_i for all periods are prescribed, where productive manhours is not to be identified with employed manhours. Two kinds of cost will be considered: (a) a wage cost for the level w_i (which takes into account the possibility of meeting the requirement r_i by use of overtime or part-time employment) and (b) a cost of changing the employment level the amount $w_i - w_{i-1}$ between successive periods. The problem is to determine an employment schedule which minimizes the sum of these costs over the entire planning interval.

The second kind of cost is the simpler to describe. Let

α = cost of hiring per worker

β = cost of termination per worker

These costs are assumed to be known, nonnegative constants; they may be interpreted as incorporating a variety of factors—cost of training, cost of paper work, etc., according to the application at hand. Then the cost of changing the employment level from w_{i-1} to w_i is

$$\alpha(w_i - w_{i-1}) \quad \text{for} \quad w_i \geq w_{i-1},$$

$$\beta(w_{i-1} - w_i) \quad \text{for} \quad w_i \leq w_{i-1},$$

or, alternatively,

$$\max \left\{ \alpha(w_i - w_{i-1}), -\beta(w_i - w_{i-1}) \right\}.$$

*Manuscript received May 7, 1957

We turn now to the first kind of cost. Suppose the productive manhour requirement r for a period is given. There will be a minimum employment level $w^*(r)$, corresponding to full use of maximum overtime necessary to meet this requirement, and for this level there will be a certain wage cost. As w increases from this minimum level, the requirement r may be met by reducing the number of employed hours; furthermore, this reduction will decrease the wage cost, since it will reduce the cost of overtime. Now suppose that there is a minimum number of employed hours guaranteed a worker, and that w' is the number of workers just necessary to meet the requirement r when all are working this minimum number of hours. As w increases beyond w', the potential productive manhours of the work force exceeds the requirement r, and the wage cost begins increasing.

These considerations lead to the assumption of a cost function

$$C(r, w)$$

and a minimum level function

$$w^*(r) ,$$

with the following interpretation and properties. The function C is defined for pairs r, w, satisfying

(1) $$w^*(r) \leq w .$$

This inequality signifies that the requirement r in any period can be met by the work force w. The value C(r, w) is the cost of the employment level w in any period when the requirement in that period is r. For fixed r, C(r, w) is initially decreasing and then increasing as a function of w; i.e., in the notation we shall adopt later, C(r, w) is a ditonic function of w for each r.

It is useful to illustrate more specifically the function C(r, w). Let a unit period be one week. Let

h = number of employed hours of a worker in one period.

Suppose that h = 16 (two days) is the guaranteed minimum h and h = 56 (seven days) is the maximum h. For simplicity, suppose that the employed manhours

$$h \cdot w$$

is the same as the productive manhours. Then the relation between h, w, and the productive manhour requirement r is graphed in Figure 1, where the cases r = 5600 and r = 11,200 are shown. The curved portion is the graph of

(2) $$hw = r ,$$

with the dotted extensions to be excluded because of the restriction $16 \leq h \leq 56$.

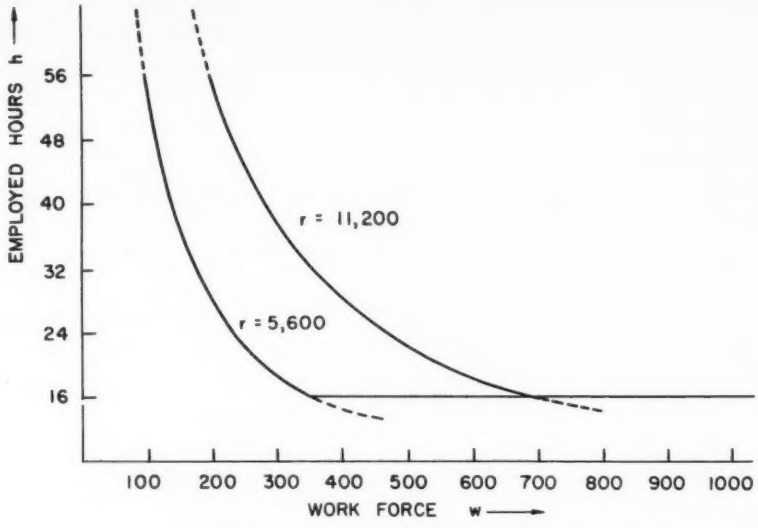


Figure 1 - Employed hours, h, of individual versus work force, w (for given productive manhour requirement, r)

Suppose that an individual worker is paid according to the scale shown in Figure 2. Eliminating h between Figures 1 and 2 yields the cost $C(r, w)$, as shown in Figure 3, where the additional case $r = 8960$ has been included. The minimum employment level $w^*(r)$ is determined by (2) with $h = 56$,

$$w^*(r) = \frac{r}{56} .$$

The foregoing example leads to a function $C(r, w)$ which is convex in w. In this paper, however, we shall assume only that C is a ditonic function of w, thereby enlarging the degree of applicability of the model. To illustrate a situation in which convexity might be too restrictive an assumption, consider the introduction of a realization or efficiency factor θ in (2), converting employed manhours to productive manhours,

$$\theta\, hw = r .$$

If θ is a constant, the situation is essentially the same as (2); but if θ is allowed to depend upon h, and possibly w, we have

$$\theta(h, w)\, hw = r .$$

More compactly, h, w, r may be related in a general manner,

$$G(h, w) = r .$$

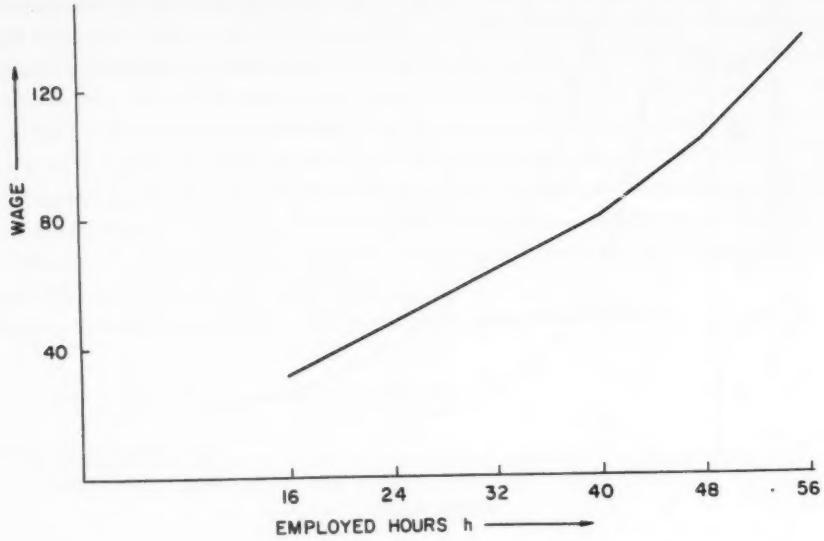


Figure 2 - Individual wage vs employed hours, h

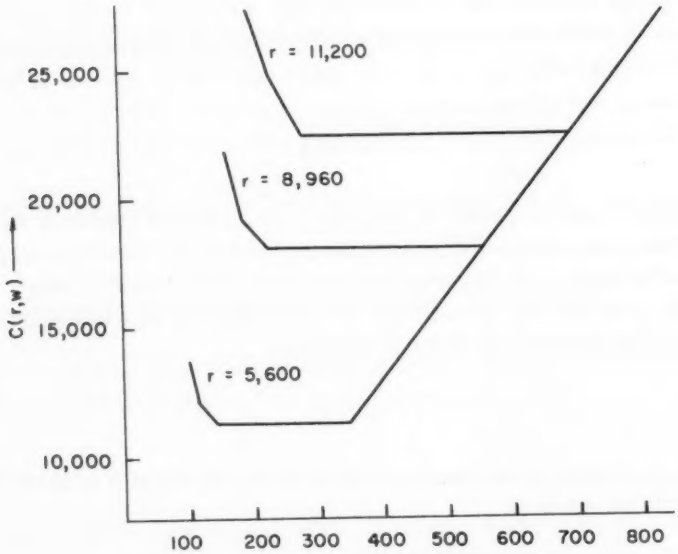


Figure 3 - Wage cost as a function of work force, w
(for given productive manhour requirement, r)

The possibility of this more general type of relationship suggests the desirability of the weaker restriction on the cost function C.

The scheduling problem may be summarized as follows. Let the requirements

$$r_1, \ldots, r_N$$

and the initial employment level $w_0 = c$, be given. Then a schedule

$$w_0, w_1, \ldots, w_N$$

is feasible in case

$$c = w_0$$

$$w^*(r_i) \leq w_i \quad , \quad i = 1, 2, \ldots, N .$$

(We may also impose an upper restriction, $w_i \leq b_i$.) To a feasible schedule assign a total cost

$$\sum_{i=1}^{N} \left[F_i(w_i) + H(w_i - w_{i-1}) \right] , \tag{3}$$

where

$$F_i(w) = C(r_i, w) \tag{4}$$

and

$$H(\Delta w) = \max \{ \alpha \cdot \Delta w, -\beta \cdot \Delta w \} \quad , \quad \alpha \geq 0, \beta \geq 0, \alpha + \beta > 0 . \tag{5}$$

Determine an optimal schedule w_i, i.e., a feasible schedule which minimizes the total cost (3).[1]

In this form, the problem may be considered one in the production planning of a non-storable commodity; since the commodity is perishable, the usual storage or inventory cost is absent, while the production cost in period i is represented by F_i and the change of production cost (independent of time) by H. A particularly interesting aspect of the problem is the nature of the production cost function (4). The mathematical model is appropriate to service commodities other than labor.

The method developed here for obtaining an optimal schedule is, in brief, the following. The given problem, in slightly modified form, is designated as Problem I. It is shown that the peak and trough values of an optimal program $\overline{w_i}$ (as graphed against the time i) occur at certain periods k_j; the precise result is contained in the "imitation" Lemma 2. The problem is then simplified to finding the peak and trough values; this is referred to as Problem II, and its exact relationship to Problem I is set forth in Theorem 2. An iterative procedure for the solution of Problem II is then given in Theorem 5, based on the approach of dynamic programming, a method developed extensively by R. Bellman [1].

[1] This formulation is mathematically adequate but implicit in it is the assumption of a rule which specifies for each r and w how the hours of employment are to be distributed among the w workers. This rule is involved in the determination of the function C(r, w), as we saw above.

It may be remarked that when the functions $F_i(w)$ are piecewise linear and convex, the problem can be elaborated into a problem in linear programming by one device or another. The present method, however, takes advantage of the special form of the minimization problem to obtain an effective algorithm which incorporates the linear as well as the essentially nonlinear case when the functions F_i are not convex.

In an earlier paper [4], we presented a method of solution for the case when the functions are all increasing (or all decreasing) and convex, a problem that rose out of a paper of Bellman, Glicksberg, and Gross [2]. In this case of monotone convex functions, the algorithm for constructing an optimal solution is somewhat simpler than the one developed in this paper, but it does not carry over to the more general problem treated here.

It is useful to know when an optimizing problem over N periods reduces to independent suboptimizing problems over fewer periods. In their interesting study of an inventory-production model [5], Modigliani and Hohn show how the method of horizon planning leads to such reduction. For the present problem we also give a criterion for the reduction of the range of the planning period, in Theorem 3. There is an interesting difference between the two results. In the case of horizon planning, it is necessary to inspect the full set of requirements through the last period N, in order to determine the suboptimum; e.g., if the requirement is altered in any period of the planning interval, or if additional periods $N+1, N+2, \ldots\ldots$ are added to the interval, then the supposed suboptimum must be reinspected for possible alteration. On the other hand, the criterion stated here depends only on the data in a certain limited range, and modifications of the requirements outside this range do not affect the determined subinterval of optimization.

The present model can be extended to several labor classes under the assumptions that (a) the classes are distinguished by different wage rates but an arbitrary worker may be assigned to any one of the classes, (b) the cost of hiring and discharging is the same for workers in all classes. Consider, e.g., employment levels u_i, v_i, respectively, for two labor classes. Let the requirements be

$$u_i^* \leq u_i \quad , \quad v_i^* \leq v_i \tag{6}$$

and let the wage cost functions be

$$D_i(u_i) \quad , \quad E_i(v_i)$$

respectively. Then the total labor force is

$$w_i = u_i + v_i \quad , \quad u_i^* + v_i^* \leq w_i \, ,$$

and the wage cost and change of employment level cost is

$$D_i(u_i) + E_i(v_i) + H(w_i - w_{i-1}) \, .$$

It is clear that for a given w_i, the u_i and v_i should be chosen to minimize the first two terms. That is, we choose F_i according to

(7) $$F_i(w_i) = \min_{u_i + v_i = w_i} \left[D_i(u_i) + E_i(v_i) \right]$$

with (6) holding. The w_i are then chosen to minimize (3) as before. It needs only to be remarked that the construction of F_i from D_i and E_i according to (7) is not difficult, and that F_i is ditonic when D_i and E_i are. Finally, reference should be made to the interesting approach of Holt, Modigliani, and Simon in their study of production and employment scheduling [6].

2. DEFINITIONS

In this section we shall list the mathematical notations and definitions which appear in the forthcoming sections. This will facilitate comprehension of the principal results for the reader who does not wish to delve into complete mathematical arguments. These results are presented later in the form of lemmas and theorems for ready reference. At the end of this section, the steps for constructing an optimum solution will be outlined.

We modify the earlier notation as follows. A program consists of $N+1$ components

$$\{x_i\} : \quad x_0, x_1, x_2, \ldots, x_N .$$

A program is feasible in case it satisfies

(8) $$a_i \le x_i \le b_i \quad , \quad i = 0, 1, 2, \ldots, N$$

where the a_i, b_i are constants. (Observe that the earlier requirement w_0 = constant, is included by taking $a_0 = b_0$. Also, for convenience, we shall treat the a_i and b_i as finite, but it is allowable that, for example, a particular b_i be $+\infty$; in that case the weak inequality is to be replaced by the strict inequality, of course.) It is convenient to picture a program as in Figure 4.

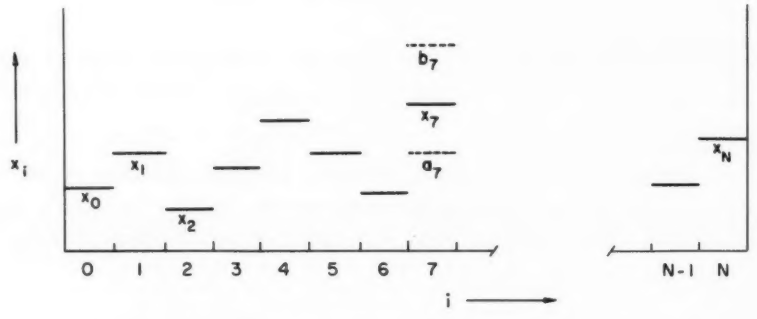


Figure 4 - Graphical representation of a program, x_i

Feasibility requires that these levels lie between the assigned levels a_i and b_i. Let

$$C(x_0, x_1, \ldots, x_n) = \sum_{i=0}^{N} F_i(x_i) + \sum_{i=1}^{N} H(x_i - x_{i-1}) , \tag{9}$$

where H is the function given by (5) and F_i is an arbitrary continuous ditonic function on the interval $a_i \le x \le b_i$. A function F is said to be ditonic[2] on $a \le x \le b$ in case there is a number m, $a \le m \le b$, such that

$$F(x) \text{ is decreasing on } a \le x \le m ,$$
$$F(x) \text{ is increasing on } m \le x \le b .$$

(Decreasing ≡ nonincreasing, increasing ≡ nondecreasing.) The number m may be a or b; thus, e.g., an increasing function on $a \le x \le b$ is ditonic on that interval. Also, a convex function is ditonic.

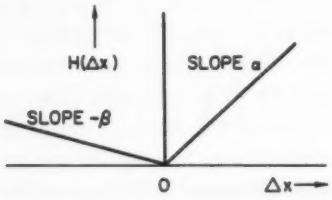


Figure 5 - The function H

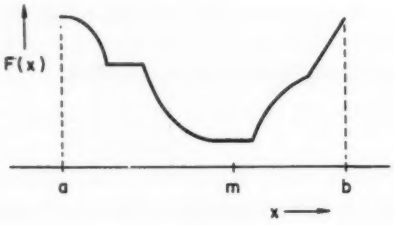


Figure 6 - Ditonic function F

Let

$$\gamma = \alpha + \beta > 0 ,$$

and let δ be a fixed number such that

$$\delta > \gamma .$$

The function F_i is defined on $a_i \le x \le b_i$; we extend it linearly with slope $-\delta$ on $x \le a_i$ and linearly with slope δ on $x \ge b_i$ to be continuous ditonic on $-\infty < x < \infty$. (If F_i is already convex, take a slope $\ge \delta$ which will preserve convexity for the extended function.)

Let

$$A = \min C(x_0, x_1, \ldots, x_N) \quad , \quad \text{subject to (8).}$$

[2]Professor T. S. Motzkin was kind enough to point out the close connection between convex functions and ditonic functions. For example, F is continuous and strictly ditonic (no horizontal segments) on $a \le x \le b$ in case F(x) has the form G(f(x)) where G is continuous convex and f is a one-to-one continuous mapping of the interval $a \le x \le b$ onto itself.

Denote by

$$\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_N$$

an optimal program, i.e., a feasible program such that

$$A = C(\bar{x}_0, \bar{x}_1, \ldots, \bar{x}_N) .$$

Let

(10) $$m_0, m_1, m_2, \ldots, m_N$$

be a feasible program consisting of minimizing values for the successive F_i

(11) $$F_i(m_i) = \min_{a_i \le x \le b_i} F_i(x) .$$

If F_i has a minimum segment, then m_i is not unique; we shall suppose a definite choice has been made.[3] The sequence $\{m_i\}$ can be partitioned into n monotonic sections:

(12)
$$\begin{cases} m_{k_0} = m_0 \le m_1 \le \ldots \le m_{k_1}, & (\text{or} \ge) \\ \quad \vdots & \vdots \\ m_{k_1} \ge m_{k_1+1} \ge \ldots \ge m_{k_2} & \vdots \\ \quad \vdots & \vdots \\ \quad \vdots & \vdots \\ \quad \vdots & \vdots \\ m_{k_{n-1}} \le m_{k_{n-1}+1} \le \ldots \le m_{k_n} \equiv m_N . & (\text{or} \ge) \end{cases}$$

(The initial, or terminal, inequalities may be of the $\ge$ type, of course, as indicated.) We suppose that a specific chain of inequalities (12) has been fixed. The integer n and the indices

$$k_j , \qquad j = 0, 1, 2, \ldots, n$$

are specified by (12). A program $\{x_i\}$ is said to <u>imitate</u> $\{m_i\}$ in case the x_i satisfy the same inequalities (12) as the m_i.

Let

$$t_j = x_{k_j} \qquad (j = 0, 1, 2, \ldots, n) .$$

---

[3] The m_i may be chosen, e.g., to yield the fewest number of monotonic sections in (12).

If $\{x_i\}$ imitates $\{m_i\}$, then the t_j satisfy alternating inequalities; e.g.,

$$t_0 \leq t_1 , \quad t_1 \geq t_2 , \quad t_2 \leq t_3 , \quad \cdots , \quad t_{n-1} \leq t_n . \tag{13}$$

Let

$$I\left[t_0, t_1, \cdots, t_n\right]$$

denote the appropriate set of alternating inequalities. We shall say that the index j is high (low) in case t_j is a relative maximum (minimum) in (13). Also the symbol

$$I\left[t_p, t_{p+1}, \cdots, t_q\right]$$

will denote that subset of inequalities of $I\left[t_0, t_1, \cdots, t_n\right]$ which involves just the variables $t_p, t_{p+1}, \cdots, t_q$. For example, for the particular inequalities exhibited in (13), 0 is low, 1 is high, n is high; and if j is high, then $I\left[t_{j-1}, t_j, t_{j+1}\right]$ stands for

$$t_{j-1} \leq t_j , \quad t_j \geq t_{j+1}$$

Let J(y) be any continuous function on $-\infty < y < \infty$ which is ultimately strictly increasing for $y \to \pm\infty$ (but not necessarily ditonic). The increasing lower envelope is defined as

$$J^+(x) = \min_{x \leq y} J(y)$$

and the decreasing lower envelope as

$$J^-(x) = \min_{y \leq x} J(y)$$

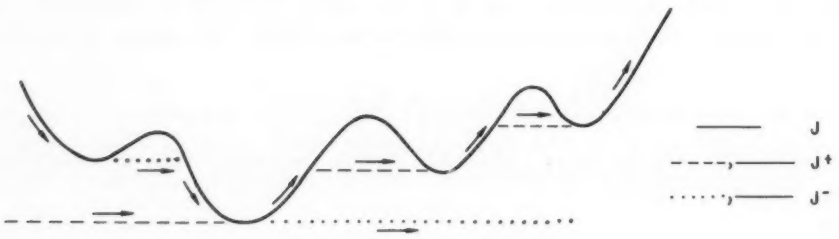


Figure 7 - The functions J, J^+, J^-

For the case of a ditonic function F with minimizing value $x = m$, the definitions yield simply

$$F^{+}(x) = \begin{cases} F(x) & \text{for } x \geq m \\ F(m) & \text{for } x \leq m , \end{cases}$$

$$F^{-}(x) = \begin{cases} F(x) & \text{for } x \leq m \\ F(m) & \text{for } x \geq m . \end{cases}$$

Given a number x, we shall mean by its right image along J^{+} the least y such that $J(y) = J^{+}(x)$, $y \geq x$. That is, if $J^{+}(x)$ coincides with $J(x)$ then we take $y = x$, while if $J^{+}(x)$ lies along a horizontal segment of J^{+}, then we slide to the right until we hit J. Similarly, we define the left image of x along J^{-} as the greatest y such that

$$J(y) = J^{-}(x) \quad , \quad y \leq x .$$

With $\{t_j\}$ subject to the restraints $I\left[t_0, t_1, \ldots, t_n\right]$, let

$$D(t_0, t_1, \ldots, t_n) = \sum_{j=0}^{n} \Phi_j(t_j)$$

where

$$\Phi_j(t_j) = F_{k_j}(t_j) \pm \gamma\, t_j + \sum_{i=k_{j-1}+1}^{k_j - 1} \left[F_i^{\mp}(t_j) - F_i(m_i)\right] + \sum_{i=k_j+1}^{k_{j+1}-1} F_i^{\mp}(t_j)$$

upper (lower) sign for j high (low) , $j = 1, 2, \ldots, n-1$.

(Recall the extension of the definition of F_i.) For $j=0$ and $j=n$ this definition is slightly modified. For $j=0$, read $+\beta$ for $+\gamma$, $-\alpha$ for $-\gamma$, and omit the first sum. For $j=n$, read $+\alpha$ for $+\gamma$, $-\beta$ for $-\gamma$, and omit the last sum. Let

$$\mu_0, \mu_1, \ldots, \mu_n$$

be defined by

$$\Phi_j(\mu_j) = \min_t \Phi_j(t) \quad , \qquad j = 0, 1, \ldots, n .$$

Two auxiliary minimizing problems will be considered.

Problem I: minimize $C(x_0, x_1, \ldots, x_N)$, $\{x_i\}$ unrestricted.

Problem II: minimize $D(t_0, t_1, \ldots, t_n)$, $\{t_j\}$ subject to $I\left[t_0, t_1, \ldots, t_n\right]$.

Let the functions $D_0, D_1, \ldots, D_n$ be defined inductively as follows:

$$D_0(t_0) = \Phi_0(t_0)$$

$$D_1(t_1) = \Phi_1(t_1) + D_0^{\mp}(t_1) \qquad \text{upper (lower) sign for } j=1 \text{ high (low)}$$

$$D_2(t_2) = \Phi_2(t_2) + D_1^{\pm}(t_2)$$

$$\vdots$$

$$D_n(t_n) = \Phi_n(t_n) + D_{n-1}^{(\mp)}(t_n) .$$

Also, let

$$\bar{t}_n \quad \text{minimize } D_n$$

$$\bar{t}_{n-1} = \text{left image of } \bar{t}_n \text{ along } D_{n-1}^- \qquad (\text{for } j=n \text{ high})$$

$$\bar{t}_{n-2} = \text{right image of } \bar{t}_{n-1} \text{ along } D_{n-2}^+$$

$$\vdots$$

$$\bar{t}_0 = \text{left image of } \bar{t}_1 \text{ along } D_0^- \qquad (\text{for } j=0 \text{ low}) .$$

For $j=n$ low, begin with "right image along D_{n-1}^+"; for $j=0$ high, end with "right image along D_0^+."

The steps for obtaining $\{\bar{x}_i\}$ may now be specified.

1. Determine minimizing values $\{m_i\}$, monotonic sections (12), and the n-1 indices k_j, $j = 1, 2, \ldots, n-1$.
2. Form the functions Φ_j.
3. Form the functions D_j and the appropriate envelopes $D_j^{\dagger}$.
4. Determine the $\bar{t}_j$.
5. Determine the $\bar{x}_i$ according to

$$\bar{x}_{k_j} = \bar{t}_j \quad , \quad j = 0, 1, \ldots, n$$

$$\bar{x}_i = \text{value closest to } m_i \text{ in the interval between } \bar{t}_j \text{ and } \bar{t}_{j+1} \text{ for } k_j < i < k_{j+1} .$$

After step 2, the criterion of Theorem 3 for reduction to a subproblem may be applied in the following manner:

a. Determine minimizing values μ_j.

b. Let p be the first index such that $I\left[\mu_{p-1}, \mu_p, \mu_{p+1}\right]$ holds.

c. Partition Problem II into two independent subproblems, one over the range $j = 0, 1, \ldots, p$, the other over $j = p, p+1, \ldots, n$.

d. Solve the first subproblem II by going on to step 3 above (with n as p, and $\bar{t}_n$ as μ_p).

e. Return to step 2 above for the second subproblem II, forming the new set of functions Φ_j^* according to

$$\Phi_0^*(t) = \begin{cases} \Phi_p(\mu_p) + \delta(t - \mu_p) & \text{for } t \geq \mu_p \\ \Phi_p(\mu_p) - \delta(t - \mu_p) & \text{for } t \leq \mu_p \end{cases}$$

$$\Phi_j^*(t) = \Phi_{p+j}(t) \quad , \qquad j = 1, 2, \ldots, n-p \, .$$

Continue the cycle of steps a - e until no further reduction is realized. (Observe that the reduction criterion of Theorem 4 is also available under special monotonicity conditions.)

3. PROBLEM I AND THE IMITATION LEMMA

The given problem is to minimize (9) subject to (8). The first step is to verify that this is equivalent to Problem I.

THEOREM 1: The program $\{\bar{x}_i\}$ is optimal for the original problem if, and only if, it is optimal for Problem I.

The proof of Theorem 1 is simplified by first noting the following lemma.

LEMMA 1: Let

$$g(x) = H(x - x_{i-1}) + H(x_{i+1} - x) \quad , \quad (x_{i-1}, x_{i+1} \text{ fixed})$$

$$c = \min \{x_{i-1}, x_{i+1}\} \, , \quad d = \max \{x_{i-1}, x_{i+1}\} \, .$$

Then g is piecewise linear with derivative

$$\frac{dg(x)}{dx} = \begin{cases} -\gamma \, , & x < c \\ 0 \, , & c < x < d \\ \gamma \, , & d < x \end{cases} .$$

The lemma is immediate from the definitions of H and γ. Observe that the lemma also applies when $i = 0$ or N, in which case, $g(x) = H(x_1 - x)$ or $g(x) = H(x - x_{N-1})$, and the conclusion is modified in an obvious way.

To prove Theorem 1, let $\{x_i\}$ be any program, not necessarily satisfying (8). With this program associate the feasible program $\{y_i\}$ given by

$$y_i = \begin{cases} a_i & \text{if } x_i \le a_i \\ x_i & \text{if } a_i \le x_i \le b_i \\ b_i & \text{if } b_i \le x_i \end{cases}$$

To prove the theorem it is sufficient to argue that if $\{x_i\}$ does not satisfy (8), then

$$C(y_0, y_1, \ldots, y_N) < C(x_0, x_1, \ldots, x_N) .$$

Hold all components x_j fixed except a particular one, $x_i = x$; then the cost function C is constant except for the terms

$$F_i(x) + g(x) ,$$

with g as in Lemma 1 (modified appropriately for an end component x_i). By the lemma and the manner of extending F_i, this sum has a strictly positive derivative for $x < a_i$ and a strictly negative derivative for $x > b_i$. Thus increasing x to a_i in the first case, or decreasing x to b_i in the second, will strictly decrease C. This completes the argument.

Henceforth, we deal with Problem I. An essential lemma to the present method is the following "imitation" lemma.

LEMMA 2: For every program $\{y_i\}$ there is a program $\{x_i\}$ which imitates $\{m_i\}$ and which does not increase C, i.e.,

$$C(x_0, x_1, \ldots, x_N) \le C(y_0, y_1, \ldots, y_N) .$$

To make the proof, we shall suppose for definiteness that the inequalities for $\{m_i\}$ begin with $\le$ (rather than $\ge$), as exhibited in (12). Number these inequalities from 1 to N in the order of increasing index.

Let $\{x_i\}$ be an arbitrary program; the proof will be made by induction on q, the number of successive inequalities (12) satisfied by the program. Consider $q = 1$. If $x_0 \le x_1$, the first step of the induction is already satisfied. Thus, suppose $x_0 > x_1$. We have $m_0 \le m_1$. Case (i), $x_0 \le m_1$. In this case, raising $x_1 = x_0$ does not increase $F_1(x_1)$ since F_1 is ditonic and $F_1(m_1) = \min$. Also, this case does not increase the second sum in (9) by Lemma 1. Hence, $x_0 \le x_1$ (in fact, $x_0 = x_1$) may be achieved without increasing C. Case (ii), $x_0 > m_1$. We may suppose $m_1 \le x_1$, for otherwise, by an argument like the one just used, we may raise x_1 up to m_1. Thus, we may assume $m_0 \le m_1 \le x_1 \le x_0$. Now decreasing x_1 to x_0 does not increase C and attains $x_0 \le x_1$. This completes the first step of the induction.

Next, suppose that $x_0, x_1, \ldots, x_q$ satisfies the first q inequalities (12), $1 \le q < N$. We shall extend q to $q+1$ by modification of the program. Case (A), $m_q \le m_{q+1}$. This inequality belongs to a chain beginning with m_p,

$$m_p \le m_{p+1} \le \cdots \le m_q \le m_{q+1} \quad , \quad p \le q .$$

By the induction assumption,

$$x_p \le x_{p+1} \le \cdots \le x_q .$$

If $x_q \le x_{q+1}$, the inductive step is complete. Suppose $x_q > x_{q+1}$. Subcase (i), $x_q \le m_{q+1}$. In this case, raising x_{q+1} to x_q yields the needed inequality. Subcase (ii), $x_q > m_{q+1}$. By raising x_{q+1}, if necessary, we may suppose $m_{q+1} \le x_{q+1}$. Thus,

$$x_q > x_{q+1} \ge m_{q+1} .$$

Let

$$s = \text{least index } p,\ p+1,\ \ldots,\ q$$

such that

$$x_{q+1} \le x_s \le x_{s+1} \le \cdots \le x_q .$$

Modify $\{x_i\}$ by lowering every component from index s through q to the common value x_{q+1}. Under this modification, the first $q+1$ inequalities (12) hold; it is required to verify that the cost C is not increased, however. Notice first that F_i is not increased for each of the modified components since $m_i \le m_{q+1} \le x_{q+1}$. Second, to treat the second sum of (9), suppose that $p \neq 0$. Then x_p is a relative "low" component. Before the change, the second sum of (9) between the fixed components x_{p-1}, x_{q+1} has the value

$$\beta(x_{p-1} - x_p) + \alpha(x_q - x_p) + \beta(x_q - x_{q+1})$$

$$= \beta(x_{p-1} - x_{q+1}) + (\alpha + \beta)(x_q - x_p) .$$

After the change, we have

$$\beta(x_{p-1} - x_p) + \alpha(x_{q+1} - x_p) \quad , \quad \text{if } s > p$$

$$\beta(x_{p-1} - x_{q+1}) \quad , \quad \text{if } s = p .$$

In either case, there is no increase. If $p = 0$, the original value is

$$\alpha(x_q - x_p) + \beta(x_q - x_{q+1}) ,$$

while after the change it is

$$\alpha(x_{q+1} - x_p) \quad , \qquad \text{if } s > p$$

$$0 \, , \qquad \text{if } s = p \, .$$

This completes the induction from q to q+1 for Case (A). Case (B), $m_q \geq m_{q+1}$, is argued analogously. This completes the proof of Lemma 2.

4. PROBLEM II

From the imitation Lemma 2, it follows we may confine ourselves to programs satisfying (12). The problem will now be reduced to n+1 dimensions in terms of the variables t_j.

For definiteness, suppose as in (13) that 0 is low and n is high. The cost function may be written

$$C = F_0(t_0) + \sum_{i=1}^{k_1-1} F_i(x_i) + F_{k_1}(t_1) + \ldots + \sum_{i=k_{n-1}+1}^{N-1} F_i(x_i) + F_{k_n}(t_n)$$

$$+ \alpha(t_1 - t_0) + \beta(t_1 - t_2) + \ldots + \alpha(t_n - t_{n-1}) \, .$$

The minimum A has the form,

$$A = \min_{I[t_0, t_1, \ldots, t_n]} \Big[F_0(t_0) - \alpha t_0 + M_1(t_0, t_1) + F_{k_1}(t_1) + \gamma t_1$$

$$+ M_2(t_1, t_2) + \ldots + F_{k_{n-1}}(t_{n-1}) - \gamma t_{n-1} \tag{14}$$

$$+ M_n(t_{n-1}, t_n) + F_{k_n}(t_n) + \alpha t_n \Big]$$

where

$$M_j(t_{j-1}, t_j) = \min \sum_{i=k_{j-1}+1}^{k_j-1} F_i(x_i) \tag{15}$$

subject to

$$t_{j-1} \leq x_{k_{j-1}} \leq \ldots \leq x_{k_j-1} \leq t_j \quad , \qquad j \text{ high}$$

with "$\geq$" in place of "$\leq$" for j low.

In the form (14), the minimizing problem lacks the property of separated variables, as it appears in Problem II. The next step is an examination of the functions M_j. The required

results are stated in the auxiliary Lemmas 3 and 4 below. In these lemmas, the functions g are understood to be continuous ditonic functions on $-\infty < x < +\infty$ with finite minimizing value v,

$$g(v) = \min g(x) .$$

Let

$$G(P, Q) = \min_{P \le x \le Q} [g(x)] . \tag{16}$$

LEMMA 3: For all $P \le Q$,

$$G(P, Q) = g^{+}(P) + g^{-}(Q) - g(v) .$$

This result follows from the ditonicity of g by considering the cases when v is in the interval $P \le x \le Q$ and when v is outside this interval.

Now consider a function of the type (15),

$$M(P, Q) = \min_{P \le x_1 \le \ldots \le x_s \le Q} \left[g_1(x_1) + \ldots + g_s(x_s)\right] .$$

LEMMA 4: Suppose that minimizing values v_j of g_j satisfy

$$v_1 \le v_2 \le \ldots \le v_s .$$

Then[4]

$$M(P, Q) = \sum_{j=1}^{s} \left[g_j^{+}(P) + g_j^{-}(Q) - g_j(v_j)\right] .$$

Furthermore, a minimizing point $\bar{x}_1, \bar{x}_2, \ldots, \bar{x}_s$ is given by

$$\bar{x}_j = \begin{cases} P & \text{if } v_j \le P \\ v_j & \text{if } P \le v_j \le Q \\ Q & \text{if } Q \le v_j \end{cases}$$

The proof is brief. The $\bar{x}_j$ defined in the lemma satisfy the necessary inequality restrictions as well as

[4] See [3] for properties of M(P, Q) without the assumption of increasing v_j but under the assumption of convex g_j.

$$g_j(\bar{x}_j) = \min_{P \le x \le Q} \left[g_j(x) \right] = G_j(P, Q) .$$

By the definition of M(P, Q),

$$\sum_{j=1}^{s} g_j(\bar{x}_j) \ge M(P, Q) \ge \sum_{j=1}^{s} G_j(P, Q) .$$

Thus, the defined $\bar{x}_j$ are optimal and

$$M(P, Q) = \sum_{j=1}^{s} G_j(P, Q) .$$

Rewriting G_j as in Lemma 3 completes the proof.

Returning to (15) and applying Lemma 4 leads to

$$M_j(t_{j-1}, t_j) = \sum_{i=k_{j-1}+1}^{k_j-1} \left[F_i^+(t_{j-1}) + F_i^-(t_j) - F_i(m_i) \right], \quad j \text{ high}$$

$$= \sum \left[F_i^-(t_{j-1}) + F_i^+(t_j) - F_i(m_i) \right], \quad j \text{ low.}$$

Substitution into (14) leads to

$$A = \min D(t_0, t_1, \ldots, t_N) \quad , \qquad \text{subject to } I\left[t_0, t_1, \ldots, t_n \right],$$

with D as defined in Section 2.

THEOREM 2: Problem I is equivalent to Problem II. Explicitly, if $\{\bar{x}_i\}$ is optimal for Problem I, then $\{\bar{t}_j\}$ with $\bar{t}_j = \bar{x}_{k_j}$ is optimal for Problem II; conversely, if $\{\bar{t}_j\}$ is optimal for Problem II, then $\{\bar{x}_i\}$ with

$$\bar{x}_{k_j} = \bar{t}_j \quad , \qquad j = 0, 1, \ldots, n$$

$$\bar{x}_i = \text{value closest to } m_i \text{ in the interval between}$$

$$\bar{t}_j \text{ and } \bar{t}_{j+1}, \quad k_j < i < k_{j+1} .$$

is optimal for Problem I.

The theorem is an immediate consequence of Lemmas 2 and 4, and the preceding discussion.

5. CONDITION FOR DECOMPOSITION INTO SUBPROBLEMS

Theorems 1 and 2 reduce the problem to Problem II. An algorithm for Problem II will be given in the next section; before presenting it we shall consider the question suggested in the title of this section.

The functions Φ_j occurring in Problem II are readily constructed from the original functions F_i. They are defined and continuous on $-\infty < t < \infty$ but, in general, not ditonic in this range. If the functions F_i are convex so are the Φ_j, and if the functions F_i are piecewise linear so are the Φ_j. This statement follows from the fact that convexity or piecewise linearity is preserved under the operation of forming $F^{\pm}$ from F and under the addition of functions.

From the definition of Φ_j we obtain

$$j \text{ high}, \quad \Phi_j(t) \text{ is } \begin{cases} \text{strictly decreasing, for } t \le a_{k_j} (\le m_{k_j}) \\ F_{k_j}(t) + \gamma t + \text{const., for } m_{k_j} \le t , \end{cases}$$

$$j \text{ low}, \quad \Phi_j(t) \text{ is } \begin{cases} F_{k_j}(t) - \gamma t + \text{const., for } t \le m_{k_j} (\le b_{k_j}) \\ \text{strictly increasing, for } b_{k_j} \le t . \end{cases}$$

(For $j = 0$ or n the coefficient is to be appropriately modified; see the remarks following the definition of Φ_j in Section 2.) For minimizing values μ_j of Φ_j we have, therefore,

$$a_{k_j} \le \mu_j \le m_{k_j} , \qquad j \text{ high}$$

$$m_{k_j} \le \mu_j \le b_{k_j} , \qquad j \text{ low} .$$

In particular, for increasing $F_i(a_i = m_i)$ or decreasing $F_i(b_i = m_i)$ we obtain j high (low) and $F_{k_j}(x)$ increasing (decreasing) on $a_{k_j} \le x \le b_{k_j}$ implies

$$\mu_j = m_{k_j} = a_{k_j}(b_{k_j}) . \tag{17}$$

THEOREM 3: Suppose that a triple of minimizing values μ_j satisfy

$$I\left[\mu_{j-1}, \mu_j, \mu_{j+1}\right] . \tag{18}$$

Then there exists an optimal program with

$$\bar{t}_j = \mu_j .$$

Similarly, for end indices,

$$I\left[\mu_0, \quad \mu_1\right] \quad \underline{\text{implies}} \quad \bar{t}_0 = \mu_0$$

$$I\left[\mu_{n-1}, \mu_n\right] \quad \underline{\text{implies}} \quad \bar{t}_n = \mu_n .$$

Before proving this theorem, we state another theorem which refines the conclusion of Theorem 3 under an added restriction of monotonicity.

THEOREM 4: Suppose that (18) holds. If j is low (high), then $F_{k_{j\mp1}}$ increasing (decreasing) on $a_{k_{j\mp1}} \le x \le b_{k_{j\mp1}}$ implies $\bar{t}_{j\mp1} = \mu_{j\mp1}$.

Similarly, if $I[\mu_0, \mu_1]$ and $j = 0$ is low (high), then this implication holds for $j + 1 = 1$, and if $I[\mu_{n-1}, \mu_n]$ and $J = n$ is low (high), then this implication holds for $j - 1 = n - 1$.

To prove Theorem 3, consider the case when j is high. By (18),

$$\mu_{j-1} \le \mu_j \ , \ \mu_j \ge \mu_{j+1} .$$

Let $\{\bar{t}_j\}$ be optimal. We may suppose

$$\bar{t}_{j-1} \le \mu_{j-1} \ , \ \bar{t}_{j+1} \le \mu_{j+1}$$

for otherwise we may lower $\bar{t}_{j-1}$, e.g., to μ_{j-1} without violating the alternating inequalities I and without increasing the function D. Hence

$$\bar{t}_{j-1} \le \mu_j \ , \ \bar{t}_{j+1} \le \mu_j .$$

Now take μ_j as $\bar{t}_j$. This establishes the conclusion of the theorem. The argument is similar for j low, and for the special cases $j = 0$ and $j = n$.

To prove Theorem 4, consider the case when j is low and $F_{k_{j-1}}$ increasing. Then $j - 1$ is high, and by (17) we have

$$\mu_{j-1} = m_{k_{j-1}} ,$$

as well as the fact that Φ_{j-1} is ditonic. Let $\{\bar{t}_j\}$ be optimal. By Theorem 3,

$$\bar{t}_j = \mu_j .$$

We may suppose that

$$\bar{t}_{j-1} > \mu_{j-1} \ge \mu_j ,$$

for otherwise we may raise $\bar{t}_{j-1}$ to μ_{j-1} and conclude the argument. (We may also suppose $j - 1 > 0$; if $j - 1 = 0$ simply put $\bar{t}_{j-1} = \mu_0$.) Consider the low $\bar{t}_{j-1}$. We may suppose

$$\bar{t}_{j-2} > \mu_{j-1} ,$$

for otherwise $\bar{t}_{j-1}$ could be lowered to μ_{j-1} to conclude the argument. Since Φ_{j-1} is ditonic, we may lower $\bar{t}_{j-1}$ into equality with $\bar{t}_{j-2}$, so that

$$\bar{t}_{j-2} = \bar{t}_{j-1} = t .$$

It follows that

$$\Phi_{j-2}(t) + \Phi_{j-1}(t) = \sum_{k_{j-3}+1}^{k_{j-2}-1} F_i^+(t) + \sum_{k_{j-2}+1}^{k_{j-1}+1} F_i^+(t) + \sum_{k_{j-2}+1}^{k_{j-1}-1} F_i^-(t)$$

$$+ \sum_{k_{j-1}+1}^{k_j-1} F_i^-(t) + F_{k_{j-2}}(t) + F_{k_{j-1}}(t) + \text{const.}$$

(The notation supposes $j - 2 > 0$; if $j - 2 = 0$ the argument is changed only in a minor way.) The first two sums on the right are increasing in t. The second two sums are constant for $t \geq \mu_{j-1}$ since $j - 1$ is high and $m_i \leq m_{k_{j-1}} = \mu_{j-1}$ for the indices i in question. Finally, the last two functions F_i are increasing for $t \geq \mu_{j-1}$. The upshot is that $t = t_{j-2} = t_{j-1}$ may be lowered to μ_{j-1} without increasing D. This yields the desired conclusion. Except for notation, the same argument applies to increasing $F_{k_{j+1}}$ and to the special cases $j = 0$ and $j = n$. For j high and decreasing functions the argument is similar.

The usefulness of Theorems 3 and 4 is that they provide simple criteria for fixing particular components $\bar{t}_j = \bar{x}_{k_j}$ of an optimal program. When such fixed components are known, the initial planning problem over $N + 1$ periods $i = 0, 1, 2, \ldots, N$ can be replaced by independent subproblems over the shorter ranges between these components. Notice that the criterion (18) depends only on data over the range of periods $i = k_{j-2}+1, k_{j-2}+2, \ldots, k_{j+2}-1$ (see the definition of the Φ_j); it is unaffected by preceding or following data. It may be remarked that fixing a particular component x_i at a constant c can be accomplished, e.g., by introducing in place of F_i the simple function with value $F_i(c)$ at c and extended linearly to the right and left of $x = c$ with slopes δ and $-\delta$, respectively.

6. ALGORITHM FOR SOLVING PROBLEM II

Define the functions E_q as follows:

$$E_0(t_0) = \Phi_0(t_0)$$

$$E_q(t_q) = \Phi_q(t_q) + \min \sum_{j=0}^{q-1} \Phi_j(t_j), \quad j = 0, 1, \ldots, n-1$$

where the minimum is taken relative to the restraints $I\left[t_0, t_1, \ldots, t_q\right]$. The desired minimum value A is given by

$$A = \min E_n(t_n) \quad , \quad t_n \text{ unrestricted.} \tag{19}$$

The functions E_q satisfy a simple recursion relation.

$$E_{q+1}(t_{q+1}) = \Phi_{q+1}(t_{q+1}) + \min \sum_{j=0}^{q} \Phi_j(t_j)$$

$$= \Phi_{q+1}(t_{q+1}) + \min_{\substack{I\left[t_q, t_{q+1}\right] \\ I\left[t_0, \ldots, t_q\right]}} \Phi_q(t_q) + \sum_{j=0}^{q-1} \Phi_j(t)$$

$$= \Phi_{q+1}(t_{q+1}) + \min_{I\left[t_q, t_{q+1}\right]} \left\{ \Phi_q(t_q) + \min_{I\left[t_0, \ldots, t_q\right]} \sum_{j=0}^{q-1} \Phi_j(t_j) \right\}.$$

That is,

$$E_{q+1}(t_{q+1}) = \Phi_{q+1}(t_{q+1}) + \min_{I\left[t_q, t_{q+1}\right]} E_q(t_q), \quad q = 0, 1, \ldots, n-1. \tag{20}$$

THEOREM 5: Let the functions D_q and the components $\bar{t}_j$ be defined as in Section 2. Then the program $\{\bar{t}_j\}$ is optimal for Problem II.

For definiteness in the proof, the case j = 0 low, j = n high will be treated. From (20) and the definition of $E^{\pm}$, we have

$$\begin{aligned} E_0(t_0) &= \Phi_0(t_0) \\ E_1(t_1) &= \Phi_1(t_1) + E_0^-(t_1) \\ E_2(t_2) &= \Phi_2(t_2) + E_1^+(t_2) \\ &\vdots \\ E_n(t_n) &= \Phi_n(t_n) + E_{n-1}^-(t_n). \end{aligned} \tag{21}$$

Furthermore, by the definition of left image and right image along E^- and E^+, respectively, we have

$$
\begin{aligned}
&E_1(t_1) = \Phi_1(t_1) + E_0(s_0) \quad , \; s_0 = \text{left image of } t_1 \text{ along } E_0^- \\
&E_2(t_2) = \Phi_2(t_2) + E_1(s_1) \quad , \; s_1 = \text{right image of } t_2 \text{ along } E_1^+ \\
&\qquad \vdots \\
&E_n(t_n) = \Phi_n(t_n) + E_{n-1}(s_{n-1}) \quad , \; s_{n-1} = \text{left image of } t_n \text{ along } E_{n-1}^- .
\end{aligned}
\tag{22}
$$

But the relations (21) are identical with those defining the functions D_q; hence

$$D_q = E_q \quad , \qquad q = 0, 1, 2, \ldots, n .$$

Now consider the program $\{\bar{t}_j\}$ defined in Section 2. First, it is clear by its definition that this program satisfies the alternating inequalities I. Second, by (19) and (22)

$$
\begin{aligned}
&A = E_n(\bar{t}_n) \\
&E_n(\bar{t}_n) = \Phi_n(\bar{t}_n) + E_{n-1}(\bar{t}_{n-1}) \\
&\qquad \vdots \\
&E_2(\bar{t}_2) = \Phi_2(\bar{t}_2) + E_1(\bar{t}_1) \\
&E_1(\bar{t}_1) = \Phi_1(\bar{t}_1) + \Phi_0(\bar{t}_0)
\end{aligned}
$$

Adding these equations yields

$$A = \sum_{j=0}^{n} \Phi_j(\bar{t}_j)$$

This completes the proof.

Since the form of Problem II is unaltered when the order of the indices is reversed, there is an analogous reverse algorithm. It takes the form

$$
\begin{aligned}
&D_n(t_n) = \Phi_n(t_n) \\
&D_{q-1}(t_{q-1}) = \Phi_{q-1}(t_{q-1}) + D_q^{\pm}(t_{q-1}) \quad , \quad q = n, n-1, \ldots, 1 \\
&\qquad \text{upper (lower) sign for } q \text{ high (low)}
\end{aligned}
$$

with the optimal program given as follows:

$$\bar{t}_0 \quad \text{minimizes } D_0(t_0)$$

$$\bar{t}_q = \text{right (left) image of } \bar{t}_{q-1} \text{ along } D_q^+(D_q^-) ,$$

$$q = 1, 2, \ldots, n .$$

The procedure embodied in Theorems 1, 2, and 5 for the determination of $\{x_i\}$ involves the successive construction of new functions out of the given functions F_i by means of the operations of adding functions and forming lower increasing and decreasing envelopes. If the given functions are piecewise linear, which is a satisfactory approximation in many applications, then so are the constructed functions; and the entire algorithm becomes a finite combinatorial procedure which may be coded for automatic computation. If the given functions are convex, then so are the constructed functions; and the determination of lower envelopes reduces to the determination of minima of convex functions. In this case, the entire algorithm rests essentially upon the operation of determining such minima and the operation of adding functions. With no further assumption than the original one of ditonicity, however, the method is relatively simple and especially well adapted to graphical techniques.

REFERENCES

[1] R. Bellman, "The Theory of Dynamic Programming," Bull. Am. Math. Soc. 60 (1954), pp. 503-516

[2] Bellman, Glicksberg, and Gross, "The Theory of Dynamic Programming as Applied to a Smoothing Problem," Jour. Soc. Ind. Appl. Math. 2 (1954), pp. 82-88

[3] W. Karush, "On a Minimization Problem Involving Convex Functions" (to appear)

[4] W. Karush and A. Vazsonyi, "Mathematical Programming and Service Scheduling," Management Science 3 (1957), pp. 140-148

[5] F. Modigliani and F. Hohn, "Production Planning over Time and the Nature of the Expectation and Planning Horizon," Econometrica 23 (1955), pp. 46-66

[6] C. C. Holt, F. Modigliani, and H. A. Simon, "Linear Decision Rule for Production and Employment Scheduling," Management Science 2 (1955), pp. 1-30

* * *

FORMULATION OF RECURRENCE EQUATIONS FOR SHUTTLE PROCESS AND ASSEMBLY LINE[1]

Richard Bellman*
The RAND Corporation†

INTRODUCTION

This paper is devoted to the study of processes which in different versions have been variously called "assembly line," "queuing," "waiting line," "bottleneck," or "scheduling processes." These may all be subsumed under the inclusive study of the flow of men and materials in space and time.

It is clear, to begin with, that quantitative studies require quantitative models, which implies, inexorably, a mathematical model of the physical system we are analyzing. Of great importance is a systematic study of the types of mathematical systems which arise from the study of flow processes. As a start on this general program, we shall discuss here two specific processes, one arising in cargo-handling, a "shuttle" process, and the other arising in industrial production. Although the processes bear some structural similarity, the basic connection is by way of similarity of mathematical treatment.

In the classification of processes, two principal types may be distinguished immediately, those of deterministic type and those of stochastic type. A process will be called deterministic if the outcome of any particular decision or action involved in the process is completely determined by the choice of the decision or action. If, however, a choice of decision or action determines a distribution of outcomes, the process will be said to be stochastic. Which process more adequately represents reality, and which is an approximation, is a matter of individual taste and philosophy, usually pragmatic.

We shall employ a mathematical technique, the method of recurrence relations, which we shall demonstrate to be equally useful for the treatment of both deterministic and stochastic processes. For deterministic processes, the recurrence relations permit a simple computation of various significant quantities associated with the process, such as delay times at various stages, either by hand or by machine. Occasionally they yield explicit analytic expressions for these quantities. If the process is stochastic, the recurrence relations allow us to calculate the distribution of these quantities by various sampling techniques. The explicit relations we obtain enable us to replace simulation methods by Monte Carlo methods, thereby

*Manuscript received February 28, 1957

†This paper was written while the author was a visiting professor in the Department of Engineering at the University of California at Los Angeles.

[1]The research for this work was sponsored by the Office of Naval Research under Contract Nonr-233(07)

increasing the effectiveness of computing machines with limited capacities. Furthermore, the analytic expressions focus the spotlight of our attention upon the combinations of parameters which are most meaningful.

Finally, we should like to point out that the recurrence technique, in its initial phase, is applied in identical fashion to both deterministic and stochastic versions.

In this report we shall concern ourselves with an application of this technique to two mathematical models, the "link-node" model of cargo-handling [1,2] and the assembly line model [3,4,5,6]. Our object is to derive the basic recurrence relations which may be used to describe the process.

I. THREE-STAGE SHUTTLE PROCESS

1. Formulation

Let us begin with a description of a three-stage shuttle process. There are four fixed positions, A, B, C, and D, with individuals X, Y, and Z stationed at A, B, and C, respectively. The function of X is to carry an item from A to B and deliver it to Y. X cannot return to A until Y has accepted the item at B. Upon receiving the item from X, Y conveys it to C, where it is delivered to C. Y cannot return to B until Z has accepted the item. As soon as Z receives the item, he conveys it to D and returns to C immediately.

Given a set of items, S, with associated transport times for the various stages, we wish to determine the time required to convey the items from A to D and the total delays incurred at B and C.

We shall obtain expressions for these quantities which can be utilized for a discussion of both deterministic and stochastic processes.

Figure 1

Define the following quantities describing the set S:

(1) x_k = the time required for X to convey the k^{th} item from A to B.

x_k' = the time required for X to return from B to A after having delivered the k^{th} item to Y.

y_k = the time required for Y to convey the k^{th} item from B to C.

y_k' = the time required for Y to return to B from C after having delivered the k^{th} item to Z.

z_k = the time required for Z to convey the k^{th} item from C to D.

z_k' = the time required for Z to return to C from D.

If the process is deterministic, we assume that the sequences (x_k), (x'_k), (y_k), (y'_k), (z_k), (z'_k) are known. If the process is stochastic, we assume that these are sequences of random variables with known distributions. Furthermore, we assume that these are independent random variables, which means no "fatigue" or "hereditary effects."

Let us further define:

(2) $d_k(X)$ = the delay to X incurred waiting for Y at B, when delivering the k^{th} item,

$d_k(Y)$ = the delay to Y incurred waiting to receive the k^{th} from X at B,

$\Delta_k(Y)$ = the delay to Y incurred waiting for Z at C, when delivering the k^{th} item,

$\Delta_k(Z)$ = the delay to Z incurred waiting to receive the k^{th} item from Y at C.

It is clear from the description of the process that:

(3) $d_k(X)\,d_k(Y) = 0$,

$\Delta_k(X)\Delta_k(Y) = 0$,

since one factor must always be zero.

Consider the following graph:

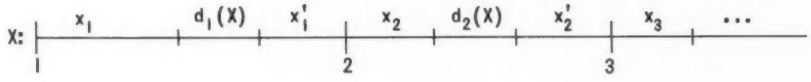


Figure 2

This is a pictorial description of the activities of X, giving the times required by X for each activity. The integers k = 1, 2, 3, ... , signal the times at which X picks up the k^{th} item at A.

Similarly we have the graph:

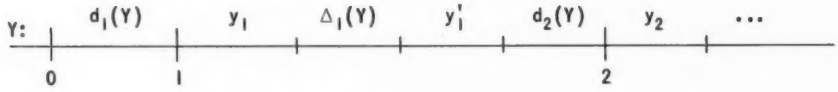


Figure 3

This describes the activities of Y. The integers, k = 1, 2, ... , signal the times at which Y picks up the k^{th} item at B.

Finally we have the graph:

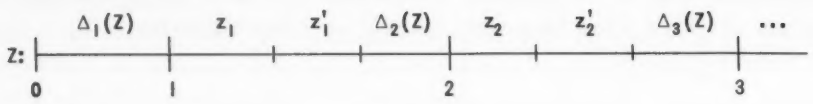


Figure 4

where the integers signal the times at which Z picks up the k^{th} item at C.

Finally, let us now define the additional quantities:

(4) r_k = the time that X arrives at B, carrying the k^{th} item,

s_k = the time that Y arrives at B, ready to receive the k^{th} item,

t_k = the time that Y arrives at C, carrying the k^{th} item,

u_k = the time that Z arrives at C, ready to receive the k^{th} item.

2. Basic Recurrence Relations

Referring to the above time scales, we have the following relations connecting the arrival times for the $(k+1)^{st}$ item with the arrival times for the k^{th}:

$$(2)\quad r_{k+1} = r_k + d_k(X) + x_k' + x_{k+1},$$

$$s_{k+1} = s_k + d_k(Y) + y_k' + y_k + \Delta_k(Y),$$

$$t_{k+1} = t_k + \Delta_k(Y) + y_k' + y_{k+1} + d_{k+1}(Y),$$

$$u_{k+1} = u_k + \Delta_k(Z) + z_k + z_k'.$$

Let us now express the delays, $d_k(X)$, $d_k(Y)$, $\Delta_k(Y)$, $\Delta_k(Z)$, in terms of these times of arrival. We have:

$$(3)\quad d_k(X) = \text{Max}\,(s_k - r_k, 0),$$

$$d_k(Y) = \text{Max}\,(r_k - s_k, 0),$$

$$\Delta_k(Y) = \text{Max}\,(u_k - t_k, 0),$$

$$\Delta_k(Z) = \text{Max}\,(t_k - u_k, 0).$$

3. A Simple Observation

A result which greatly simplifies our subsequent results is the following:

$$(1)\quad d_k(X) - d_k(Y) = \text{Max}\,(s_k - r_k, 0) - \text{Max}\,(r_k - s_k, 0)$$

$$= s_k - r_k\,,$$

with the corresponding result for $\Delta_k(Y) - \Delta_k(Z)$,

$$(2)\quad \Delta_k(Y) - \Delta_k(Z) = u_k - t_k\,.$$

4. Explicit Recurrence Relations

Let us now combine the results of the preceding two sections so as to obtain simple recurrence relations connecting $\Delta_k(Y)$ with $\Delta_{k-1}(Y)$ and $d_k(Y)$ with $d_{k-1}(Y)$.

Combining (2.2) and (3.2), we have:

$$(1)\quad u_{k+1} - t_{k+1} = u_k - t_k + \Delta_k(Z) - \Delta_k(Y)$$

$$+ (z_k + z_k' - y_k' - y_{k+1}) - d_{k+1}(Y)$$

$$= (z_k + z_k' - y_k' - y_{k+1}) - d_{k+1}(Y)\,.$$

Hence, referring to (2.3),

$$(2)\quad \Delta_{k+1}(Y) = \text{Max}\,(u_{k+1} - t_{k+1}, 0)$$

$$= \text{Max}\left[(z_k + z_k' - y_k' - y_{k+1}) - d_{k+1}(Y), 0\right].$$

Similarly,

$$(3)\quad s_{k+1} - r_{k+1} = s_k - r_k + d_k(Y) - d_k(X) + (y_k + y_k' - (x_k' + x_{k+1}))$$

$$+ \Delta_k(Y)$$

$$= (y_k + y_k' - (x_k' + x_{k+1})) + \Delta_k(Y)\,.$$

Hence,

$$(4)\quad d_{k+1}(Y) = \text{Max}\,(r_{k+1} - s_{k+1}, 0)$$

$$= \text{Max}\left[-\Delta_k(Y) - (y_k + y_k' - x_k' - x_{k+1}), 0\right].$$

Combining (2) and (4), we have:

$$(5)\quad \Delta_{k+1}(Y) = \operatorname{Max}\left[\alpha_k - \operatorname{Max}\left(-\Delta_k(Y) + \beta_k\,,\ 0\right),\ 0\right],$$

where we have set for simplicity of notation,

$$(6)\quad \alpha_k = z_k + z_k' - y_k' - y_{k+1}$$

$$\beta_k = y_k + y_k' - x_k' - x_{k+1}\,.$$

We can simplify (5) slightly,

$$(7)\quad \Delta_{k+1}(Y) = \operatorname{Max}\left(\alpha_k + \Delta_k(Y) - \operatorname{Max}\left(\Delta_k(Y),\ -\beta_k\right),\ 0\right),$$

which can also be written

$$(8)\quad \Delta_{k+1}(Y) = \operatorname{Max}\left(\Delta_k(Y),\ -\beta_k,\ \alpha_k + \Delta_k(Y)\right) - \operatorname{Max}\left(\Delta_k(Y),\ -\beta_k\right).$$

This is the fundamental recurrence relation, a basis for either a Monte Carlo treatment or an analytic discussion.

Referring to (4),

$$(9)\quad d_{k+1}(Y) = \operatorname{Max}\Big[-\operatorname{Max}\left(z_{k-1} + z_{k-1}' - y_{k-1}' - y_k - d_k(Y)\,,\ 0\right) - \left(y_k + y_k' - x_k' - x_{k+1}\right),\ 0\Big].$$

This can be written

$$(10)\quad d_{k+1}(Y) = \operatorname{Max}\Big[\left(-z_{k-1} - z_{k-1}' + y_{k-1}' + d_k(Y) - y_k' + x_k' + x_{k+1}\right),\ 0\,,\ -z_{k-1} - z_{k-1}' + y_{k-1}' + y_k + d_k(Y)\Big] - \operatorname{Max}\Big[-z_{k-1} - z_{k-1}' + y_{k-1}' + y_k + d_k(Y)\,,\ 0\Big].$$

or, referring to (6),

$$(11)\quad d_{k+1}(Y) = \operatorname{Max}\Big[-\alpha_{k-1} + d_k(Y) - \beta_k,\ 0,\ -\alpha_{k-1} + d_k(Y)\Big] - \operatorname{Max}\Big[-\alpha_{k-1} + d_k(Y)\,,\ 0\Big],$$

or

$$(12)\quad d_{k+1}(Y) = \operatorname{Max}\Big[-\beta_k,\ \alpha_{k-1} - d_k(Y)\,,\ 0\Big] - \operatorname{Max}\Big[\alpha_{k-1} - d_k(Y)\,,\ 0\Big].$$

This is also a fundamental recurrence relation.

II. THE GENERAL N-STAGE SHUTTLE PROCESS

1. Formulation

Although the above notation was satisfactory in describing a three-stage process, if we wish to describe a general N-stage process, we must employ a better notation. Consequently, we shall begin all over again, introducing a new notation.

Figure 5

An item arriving from 0 is picked up at P_1 by X_1, who conveys it to P_2 and waits until X_2 picks it up. X_1 then returns to P_1 to await the next item. The process is repeated all down the line, until P_N, where X_N transports the item to T, the terminal.

Let

(1) $x_k(i)$ = the time required by X_k to convey the i^{th} item from P_k to P_{k+1} after having received it from X_{k-1},

$x'_k(i)$ = the time required by X_k to return to P_k from P_{k+1} after having conveyed the i^{th} item to X_{k+1},

and let us define the delays

(2) $d_k(i)$ = the delay encountered by X_k awaiting the i^{th} item at P_k,

$\Delta_k(i)$ = the delay encountered by X_k waiting to deliver the i^{th} item at P_{k+1},

for $k = 1, 2, \ldots, N$.

Finally, we require the times of arrival,

(3) $r_k(i)$ = the time of arrival of X_k at P_k, ready to receive the i^{th} item,

$s_k(i)$ = the time of arrival of X_k at P_{k+1}, ready to deliver the i^{th} item.

2. Connection Between Delay Times and Arrival Times

We have the following relations connecting the above quantities

(1) $d_k(i) = \text{Max}\,(s_{k-1}(i) - r_k(i)\,,\, 0)\,,$ $\qquad k = 1, 2, \ldots$

$\Delta_k(i) = \text{Max}\,(r_{k+1}(i) - s_k(i)\,,\, 0)\,.$ $\qquad s_o(i) = 0$

From these, we conclude that

$$(2)\quad d_k(i) - \Delta_{k-1}(i) = s_{k-1}(i) - r_k(i) .$$

3. Basic Recurrence Relations

Let us now write down the equations connecting $r_k(i+1)$ and $s_k(i+1)$ with $r_k(i)$, $s_k(i)$, and the delays.

We have

$$(1)\quad r_k(i+1) = r_k(i) + d_k(i) + x_k(i) + x_k'(i) + \Delta_k(i) ,$$

$$s_k(i+1) = s_k(i) + \Delta_k(i) + x_k'(i) + d_k(i+1) + x_k(i+1) .$$

Thus

$$(2)\quad r_k(i+1) - s_{k-1}(i+1) = \Delta_k(i) - d_{k-1}(i+1) + x_k(i) + x_k'(i) - x_{k-1}'(i) - x_{k-1}(i+1) .$$

Hence

$$(3)\quad \Delta_{k-1}(i+1) = \operatorname{Max}\left[\Delta_k(i) - d_{k-1}(i+1) + \alpha_k(i) , 0\right],$$

where

$$(4)\quad \alpha_k(i) = x_k(i) + x_k'(i) - x_{k-1}'(i) - x_{k-1}(i+1) .$$

Similarly, since

$$(5)\quad s_{k-1}(i) - r_k(i) = d_{k-1}(i) - \Delta_k(i-1) + x_{k-1}'(i-1) + x_{k-1}(i) - x_k(i-1) - x_k'(i-1) ,$$

we have

$$(6)\quad d_k(i) = \operatorname{Max}\left[\beta_k(i) + d_{k-1}(i) - \Delta_k(i-1) , 0\right],$$

where

$$(7)\quad \beta_k(i) = x_{k-1}'(i-1) + x_{k-1}(i) - x_k(i-1) - x_k'(i-1) .$$

III. GENERAL SHUTTLE PROCESS

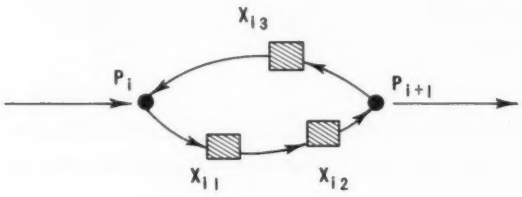
1. Storage

Let us now mention in passing a number of modifications of the above process which it may be desirable to introduce.

The first of these involves storage. Instead of assuming that X_i is required to hand each item over to X_{i+1} at P_{i+1} before returning to his position at P_i, let us assume that there are storage facilities at P_{i+1}, allowing a number of items up to k_{i+1} to be stored there until X_{i+1} can transport them. This is a very important modification, since much of the delay can be avoided with suitable storage capacity.

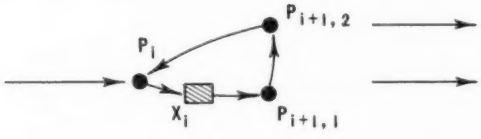
2. Multiple Shuttles

The process treated in the preceding section considered only one conveyor going back and forth between each position. A more general situation involves m_i shuttles traversing the distance between P_i and P_{i+1}. Schematically,



3. Multistop Shuttles

In some situations, the same shuttle must make a number of stops before returning to its original position. Schematically,



The shuttle, X_i, stops at $P_{i+1,1}$ and $P_{i+1,2}$, in this order, before returning to P_i.

IV. TWO-STAGE ASSEMBLY LINE

1. Formulation

Let us now turn to the consideration of a flow process of a different type. We shall discuss a simple mathematical model of a two-stage assembly line. As will be seen from the description, the model is equally useful in the consideration of a number of other waiting-line processes.

Let us suppose that we have two machines, M_1 and M_2, performing different functions, and a number of items, N, which require processing on these machines. Let

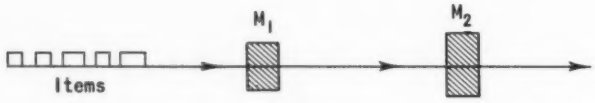
(1) x_i = the time required by the i^{th} item on the first machine,

y_i = the time required by the i^{th} item on the second machine,

and assume that each item must pass through the first machine before going through the second machine.

The process is carried through in the following way. The first item is fed into M_1, requiring a time x_1, and then sent to M_2, requiring a time y_1 on this machine. As soon as the first item has been processed by M_1, the second item is fed in. As soon as the second item has passed through this first stage, it is sent to M_2. Here there may or may not be a delay, depending upon whether M_2 has completed the processing of the first item or not. The process continues in this way.

Schematically,



The problem is to determine the total delay incurred in the processing of the N items, in the deterministic case, and the distribution of the delay times if the process is stochastic.

2. Recurrence Relations

Define the following quantities

(1) d_k = the delay incurred by the k^{th} item waiting upon the second machine,

r_k = the time of arrival of the k^{th} item at the waiting line for the second machine.

The basic recurrence relations are

$$(2)\quad d_{n+1} = \text{Max}\left[d_n + y_n - x_{n+1}, 0\right],$$

$$r_{n+1} = r_n + x_{n+1}.$$

By the use of (2), an interesting explicit formula for d_N can be derived. We have

$$(3)\quad d_1 = 0,$$

$$d_2 = \text{Max}\left[y_1 - x_2, 0\right],$$

$$d_3 = \text{Max}\left[\text{Max}\,(y_1 - x_2, 0) + y_2 - x_3, 0\right]$$

$$= \text{Max}\left[y_1 + y_2 - x_2 - x_3, y_2 - x_3, 0\right],$$

whence it follows inductively that

$$(4)\quad d_N = \operatorname{Max}\left[\operatorname*{Max}_{1\le k\le N-1}\left[\sum_{i=k}^{N-1} y_i - \sum_{i=k+1}^{N} x_i\right], 0\right]$$

The result is due to S. Johnson [4].

3. Optimal Arrangement

In the deterministic case, an interesting problem is that of determining the order in which the items should be fed into the machines so as to minimize the total time required by the process. This problem was resolved by S. Johnson [4] by using the explicit expression given in (2.4) and by the author [3], by using the functional equation technique of dynamic programming. In this latter paper, continuous versions of the process are also discussed.

Although the two-stage process can be treated in a number of ways, the three-stage process, and general N-stage process, so far have resolutely defied either an analytic or computational approach.

V. THREE-STAGE PROCESS

1. Formulation

Let us now consider a corresponding three-stage process.

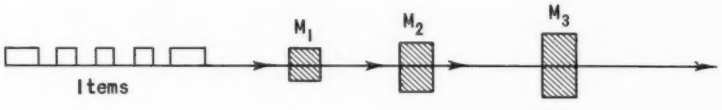
Let

(1) x_i = the time required by the i^{th} item on M_1,

y_i = the time required by the i^{th} item on M_2,

z_i = the time required by the i^{th} item on M_3.

Schematically,



Further, let

d_i = the delay incurred by the i^{th} item waiting on the second machine,

Δ_i = the delay incurred by the i^{th} item waiting on the third machine,

r_i = the time of arrival of the i^{th} item at the waiting line for the second machine,

s_i = the time of arrival of the i^{th} item at the waiting line for the third machine.

These last two times are times of arrival. The times at which the items are processed are the arrival times plus the delay times.

2. Recurrence Relations

The basic recurrence relations are

$$(1)\quad r_k = \sum_{\ell=1}^{k} x_\ell ,$$

$$s_k = r_k + d_k + y_k ,$$

$$d_k = \text{Max}\left[- r_k + (r_k + d_{k-1} + y_{k-1}) , 0 \right],$$

$$\Delta_k = \text{Max}\left[- s_k + (s_{k-1} + \Delta_{k-1} + z_{k-1}) , 0 \right],$$

These may be simplified to

$$(2)\quad d_k = \text{Max}\left[d_{k-1} + y_{k-1} - x_k , 0 \right],$$

$$\Delta_k = \text{Max}\left[\Delta_{k-1} + z_{k-1} + d_{k-1} + y_{k-1} - x_k - d_k - y_k , 0 \right].$$

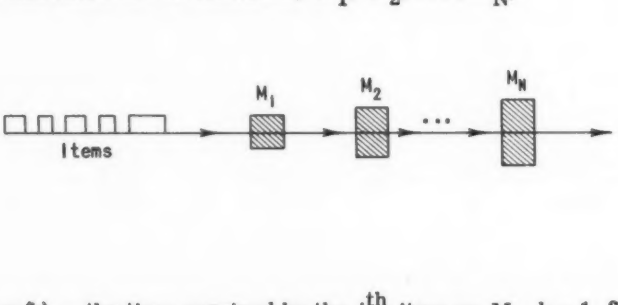
Furthermore, an explicit expression corresponding to (2.4) can be obtained (see Johnson [4]).

VI. N-STAGE PROCESS

1. Formulation

As before, we must introduce a new notation to treat the N-stage process.

We suppose that there are N machines, $M_1, M_2, \ldots, M_N$,



Let

(1) $x_i(k)$ = the time required by the i^{th} item on M_k, $k = 1, 2, \ldots, N$, $i = 1, 2, \ldots m$,

$d_i(k)$ = the delay incurred by the i^{th} item waiting on the k^{th} machine,

$r_i(k)$ = the time of arrival of the i^{th} item at the waiting line for the k^{th} machine.

2. Recurrence Relations

The basic recurrence relations are

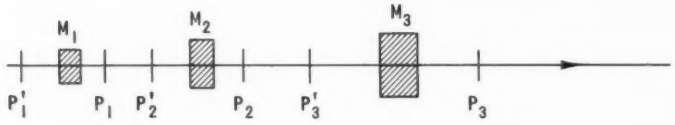
$$(1)\ (a)\quad r_i(k) = r_i(k-1) + d_i(k-1) + x_i(k-1)$$

$$(b)\quad d_i(k) = \text{Max}\left[r_{i-1}(k) + d_{i-1}(k) + x_{i-1}(k) - r_i(k)\ ,\ 0\right].$$

VII. ASSEMBLY-LINE AND SHUTTLE PROCESS

1. Discussion

A number of flow processes combine features of both the assembly-line model and the shuttle model. Consider a situation where there are a number of "machines," $M_1, M_2, \ldots, M_N$, placed in this order, and a number of items which must pass through these machines or stages. Schematically,



The item is received at P_k' and placed on the waiting line for the machine M_k. After having gone through the machine, it is placed on a waiting line for conveyance from P_k to P_{k+1}'. The process continues in this way.

We shall postpone any precise formulation of more general processes of this type until a later time, since our numerical and analytic pilot studies will center about the simpler models discussed in the foregoing pages.

ACKNOWLEDGMENTS

I would like to thank Y. Fukuda and M. Pollack, associated with the Department of Engineering of UCLA, for their careful reading of this paper, and for their many helpful corrections and comments.

BIBLIOGRAPHY

[1] "An Engineering Analysis of Cargo - Handling," Report 53-21, Dept. of Engineering, UCLA, August 1953

[2] "An Engineering Analysis of Cargo - Handling," Report 55-2, Dept. of Engineering, UCLA, November 1954

[3] R. Bellman and O. Gross, "Some Combinatorial Problems Arising in the Theory of Multi-Stage Process," Jour. Soc. Ind. Appl. Math. 2 (1954), pp. 175-183

[4] S. Johnson, "Optional Two- and Three-Stage Production Schedules with Setup Times Included," Naval Research Logistics Quarterly, March, 1954

[5] M. Salveson, "On a Qualitative Method in Production Planning and Scheduling," Econometrics 20 (1952), pp. 571-4

[6] A. Vazsonyi, "The Use of Mathematics in Production and Inventory Control," Management Science 1 (1954), pp. 70-85

An excellent discussion of queuing theory, together with an extensive bibliography, may be found in:

D. G. Kendall, "Some Problems in the Theory of Queues," Jour. Royal Stat. Soc., Ser. B, Vol. XIII (1951), pp. 151-185

A general discussion of mathematical problems arriving in flow processes may be found in:

R. Bellman, "Mathematical Aspects of Scheduling Theory," Jour. Soc. Ind. Appl. Math., Vol. 4, (1956) pp. 168-205.

where a brief bibliography of this field may be found.

* * *

RELATIONSHIPS BETWEEN WEAPONS AND LOGISTICS EXPENDITURES

Murray A. Geisler*
The Rand Corporation

INTRODUCTION

One of the objectives of the Air Force is to allocate its budget to the purchase of weapons[1] and to expenditures on their logistics support in a way that maximizes operational capability. The purpose of this paper is to discuss general problems involved in making this kind of decision.

I. OPERATIONAL READINESS AND LOGISTICS

The growing speed and destructiveness of airborne weapons has changed radically the meaning of operational capability from what it used to be in World War II. Strategic air power can no longer be measured in terms of a gradual build-up of capacity to deliver, over many months, thousands of sorties against the enemy. Instead, it must be measured in terms of the instantaneous striking power of a limited number of aircraft flying a limited number of sorties. Operational readiness is the overriding criterion.

Similarly, since the United States has become vulnerable to sudden and heavy attack, a strong air defense force, kept in a state of perpetual readiness, is needed. The operational capability of tactical air forces, too, depends on the maintenance of a high degree of alertness and mobility for possible action anywhere on the globe.

With the exception of a prolonged, limited war of the Korean type—perhaps an unlikely thing in the future—the foreseeable conditions of war, peripheral as well as total, defensive as well as offensive, emphasize the value of continued operational readiness for SAC, ADC, and a major part of TAC forces.

From this emphasis on readiness, there flow two simple requirements: a high premium on operationally ready weapons and a high premium on operationally ready crews. It is clear that aircraft not operationally ready because of lack of maintenance would contribute little to the war effort if D-day came tomorrow. They might be lost on their bases, or the war might be decided before they could enter it. Likewise, a lack of operationally ready crews capable of handling the complex new weapons would constrain the operational capability of the Air Force in the critical early stages of the war.

*Manuscript received April 16, 1957

[1] The discussion will be centered on aircraft as major weapons; but the concepts proposed here probably apply equally well to large missiles and their logistic support.

Meeting these two requirements is not as simple as the logic which leads to their definition. First, to keep up-to-date weapons in a high state of operational readiness demands a great logistics effort. Second, to make and keep crews operationally ready requires intensive training (in the flying of combat aircraft, for example), and flying is bound to make some aircraft not operationally ready because of needed maintenance or parts.[2] Thus the training requirement increases further the logistics effort needed to maintain an operationally ready force in being.

The problem outlined in this paper is the allocation of available budget funds between weapons and their logistics support so as to obtain the maximum number of operationally ready aircraft, given the flying program required for crew training. Since additional operationally ready aircraft can be obtained either by buying more aircraft or by reducing the number of aircraft not operationally ready, a problem of budget allocation does exist. Our discussion will deal only with aircraft not operationally ready because of needed parts or maintenance. Of course, there are other reasons for aircraft not being operationally ready, such as modification and technical order compliance, but these involve other considerations not as directly related to the logistics system, such as engineering design changes and poor equipment design or manufacture. Before turning to the analysis of the relation between expenditures on weapons and expenditures on their logistics support, it might be useful to consider a simple yet comprehensive outline of the logistics effort.

II. OUTLINE OF THE LOGISTICS EFFORT

A. Supply Support and Maintenance Support

Air Force logistics comprise the activities involved in holding weapons in, or restoring them to, operationally ready status. It includes the maintenance, supply, procurement, transportation, communication, disposal, etc., activities of the Air Force. Organizationally, it includes the Air Materiel Command and substantial elements of each of the other commands and of United States Air Force Headquarters.

For analysis, it is useful to simplify the picture of the logistics effort. This can be done by subdividing it into supply support and maintenance support. For present purposes, we define supply support and maintenance support so broadly as to make them encompass virtually all logistics activities.

In supply support we include all the activities that more or less directly affect the number of aircraft not operationally ready because of needed parts (AOCP) and the number of aircraft not fully equipped (ANFE). In a more direct sense, supply support includes all activities that affect the stock of parts, the procurement or repair of parts, their storage at bases and depots, and their cataloguing, distribution, warehousing, and transporting. Local manufacture of parts at bases or depots also is classified under supply support, although the work is often done in maintenance shops.

In maintenance support we include all the activities that more or less directly affect the number of aircraft not operationally ready because of lack of maintenance. These activities include periodic maintenance at base level, unscheduled maintenance, compliance with Technical Orders, as well as IRAN (inspection and repair as necessary) at depots or

[2] In the case of missiles, the training requirement does not exist in the same form as for manned aircraft; but to the extent that the missiles generate maintenance and supply requirements in peacetime, the principles in this paper apply.

contractors' facilities. Thus, maintenance support includes activities under operating commands as well as under AMC.[3]

This classification of supply support and maintenance support is based on the activity or job performed rather than on the organization that does it. Field maintenance repairs whole aircraft, and it repairs spare parts. When it does the first, it acts as part of maintenance support; and when it does the second, it acts as part of supply support. The natural interaction of "supply" and "maintenance," taken in their organizational meaning—e.g., in the form of local manufacture ("maintenance") making up for "supply" shortages, or of ample "supply" making up for lacking repair capacity—is eliminated by these definitions. Supply support in this analysis, let us repeat, is any activity supplying parts; and maintenance support is any other activity keeping aircraft in commission or restoring them to that status.

B. Logistics System Capacity

Logistics system capacity affects in two ways the rate at which aircraft not operationally ready are restored to serviceability: first, through the promptness with which each malfunction is acted upon, and secondly, through the speed with which it is relieved.[4] We shall now discuss maintenance support and supply support capacity in these terms.

If we fix the amount of flying for crew training, we determine in effect a volume of logistics work to be performed, although the precise amount cannot be foretold. Much of the demand for supply and maintenance is erratic [1]. Therefore, there is some probability that the repair and supply services demanded will exceed the given amount of capacity to provide them. When the maintenance support capacity is exceeded, we may have an aircraft awaiting maintenance and not operationally ready, and when the supply system is unable to meet the demand for an aircraft spare part, we may have an aircraft not operationally ready because of needed parts.

In addition, the capacity of the logistics system determines the speed with which an aircraft not operationally ready because of needed parts or maintenance is restored to operational readiness. It takes some time to supply the parts or to render the services needed to correct a malfunction. Thus, in the case of maintenance, the more manpower and equipment available to work on a given aircraft, the more quickly the aircraft is restored to operationally ready status. Of course, at some point the additional resources add very little to the benefits obtained, such as when a particular job can use the resources of only two men and a third man cannot be of any assistance. But within reasonable limits, the statement is valid.

---

[3] As noted before, we classify the repair of parts both at field and depot level under supply support because its immediate purpose is to add to the supply of parts.

[4] These are in turn affected by the frequency with which malfunctions have to be treated by the given logistics capacity.

[1] Brown, B. B., Geisler, M. A., Analysis of the Demand Patterns for B-47 Airframe Parts at Airbase Level, The RAND Corporation, Research Memorandum RM-1297, July 27, 1954; Geisler, M. A., Youngs, J. W. T., Predictability of Demand for B-47 Airframe Spare Items, The RAND Corporation, Research Memorandum RM-1300, August 4, 1954; LaVallee, R. S., Stoller, D. S., The Effect of Maintenance and Reliability on the Operational Effectiveness of an Interceptor Squadron, A Case Study (Project Lock-On), (U), the RAND Corporation, Research Memorandum RM-1499, June 8, 1955 (Secret).

III. THE WEAPONS/LOGISTICS BALANCE

A. The Problem

The problem is this: how can the resources available for purchasing both a weapon and its logistics support be so allocated as to obtain the maximum number of operationally ready[5] aircraft, given the flying hour program required for training.[6]

Experience indicates that a large number of aircraft are not operationally ready at all times [2]. As long as there is any flying activity this is probably inevitable; but given a certain flying-hour program, the number of aircraft not operationally ready can be reduced by devoting more resources to logistics support in the form of parts, maintenance personnel or equipment, faster communication and transportation, etc. If such an increase in logistics expenditure were to come out of a fixed over-all budget, however, a corresponding decrease in the purchase of the weapons themselves would become necessary. On the other hand, the advent of a program to buy more aircraft would decrease the budget portion available for logistics support and thus tend to raise the number of aircraft not operationally ready. Thus a given budget can be apportioned in several different ways among aircraft, supply, and maintenance; and the number of operationally ready weapons depends in part upon how this apportionment is made.

The best allocation depends on the costs of obtaining additional operationally ready aircraft, either by spending more on new aircraft or by spending more on supply support or more on maintenance support. Thus, if an additional expenditure on supply equivalent to the cost of one aircraft would tend to make two more aircraft operationally ready, this would be a cheaper way of increasing the operational-readiness level than would the purchase of one additional aircraft. Similarly, if shifting some funds from supply to maintenance would reduce the number of aircraft not operationally ready because of needed maintenance more than it would increase the number of aircraft not operationally ready because of needed parts, this too would be an economical reallocation of funds.

In general, as long as the number of operationally ready aircraft can be increased by shifting funds from one category to another—without impairing the flying accomplishment or increasing the total budget—it is desirable to make the shifts. Theoretically, an optimal allocation could be achieved in which the balance of supply, maintenance, and weapons is such that any further shift of resources from one to the other would reduce the number of operationally ready aircraft. Obviously there are great practical obstacles to making such a nice adjustment. Lack of data may obscure the relevant facts, prior decisions may limit the scope of the decision maker, etc. Nevertheless, the theoretical optimum and the adjustments capable of leading to it are described in the following section because, by understanding them, the decision maker may become aware of the shifts that are possible in a practical situation.

This problem is central to over-all weapon-system management. It presents itself in different forms during the life of the aircraft. In the early planning stages, there is full

---

[5] The term "operationally ready" has recently been adopted by the USAF in place of such other descriptions as "in commission" or "combat ready."

[6] This is, of course, only a part of the whole problem of allocating resources to support a weapon system. On another level of analysis, such things as the desirable level of crew training, the trade-off between flying hours and the use of training aids, and the balancing of civilian personnel might all be considered.

[2] Geisler, M. A., Mirkovich, A. R., The Relation of Aircraft Status Data to the Logistics System, (U), The RAND Corporation, Research Memorandum RM-1640, February 2, 1956 (Confidential).

freedom of choice among the types of resources but little reliable information about logistics costs and needs. At that time the problem is to plan a rough balance among supply, maintenance, and aircraft which promises to produce the largest average number of operationally ready aircraft throughout the life or operational service period of the aircraft. Later, when significant numbers of the aircraft are in the inventory, there is better information on logistics costs and needs; but by that time, much of the budget is likely to have been committed for aircraft and their provisioning, and the opportunities for shifting budget funds between procurement and logistics support will have become limited.

The allocation problem which has been presented in this cursory way will be treated more rigorously in Section IV below, and a solution will be outlined there in a general form. In order to proceed, some assumptions have to be made which, while restrictive and unrealistic, help in tracing the main features of the problem.

B. Assumptions

First, the best balance among the elements we are considering is apt to change during the life of a weapon, but this dynamic consideration is ignored in the present "static" treatment of the problem. In a static treatment, we take the life of the weapon as being represented by certain averages—average number of operationally ready aircraft, average cost of aircraft and support, etc. This simplification is necessary to permit a preliminary solution. Later studies may deal with the dynamics of the problem.

Second, it is assumed that supply support and maintenance support are independent of each other in the sense that adding more of one of these factors will not directly affect the productivity of the other. This assumption is imposed through the way in which we have defined supply and maintenance support. (See II.A.) The remaining problem is, of course, how those definitions can be translated into practical distinctions.

Third, changes in the use of resources within the maintenance and supply support areas are not considered at all. It is assumed that the "state of the art" in maintenance and supply support is fixed, and that one cannot get more maintenance or supply by greater efficiency. (One probably can.)

Fourth, it is assumed that the prices of the various resources used in supply and maintenance do not change as a result of shifts in the allocation of funds. Or, putting it another way, we assume that more of the same kinds of skills, parts, equipment, etc., can be purchased at the going price. In fact, this assumes away bottlenecks.

Fifth, it is assumed that as more and more resources are devoted to maintenance or supply the benefit derived from the additional units in terms of decreased nonoperational aircraft will lessen. This means that, for both supply and maintenance, the well-known assumption of diminishing returns is assumed to operate. Over the practical ranges it is probably a valid one. The assumption means that as increasingly large amounts of maintenance resources are applied, their productivity, per unit, will decline; as more are added they will make smaller and smaller additional contributions. The same is assumed with regard to supply support.

The following section outlines, in geometrical form, a method of solving the allocation problem. The reader will find in it the general principles of the proposed approach.

IV. GEOMETRIC PRESENTATION OF THE ALLOCATION METHOD

In this geometric presentation of the allocation problem we begin by looking for an optimal allocation, where all three of the elements discussed above are in balance. Such an allocation might be aimed at early in a weapon's life when there is freedom to allocate the budget in any way. In cases where not all of the information is available or where some of the budget is already committed, the same concepts may be applied to establish improved allocations. This situation will be taken up in part D of this section.

The following problem will be illustrated: Given an objective of keeping a specified number of aircraft operationally ready while meeting a flying-hour program, we wish to find the combination of aircraft, maintenance, and supply which will meet the objective at minimum budget (or cost).[7] Conceptually, there is no difference between the process used to solve this problem and that used to solve the inverse problem; i.e., given a specified budget, how should this budget be allocated among aircraft, supply, and maintenance support so as to maximize the number of operationally ready aircraft for a given flying-hour program. We have chosen to illustrate the solution to the first formulation because it seems to be a more useful statement of the long-range problem in which a military objective is given and the resources for accomplishing this objective are to be determined.

The steps involved in this solution are as follows:[8]

1. Finding the minimum logistics system cost corresponding to each aircraft cost that will meet the stipulated objective of operationally ready aircraft and flying hours.

2. Finding that combination of logistics system cost and aircraft cost which meets the objective at minimum total budget (cost of aircraft plus supply plus maintenance support).

3. Breaking down the total logistics system cost obtained in step 2 into its maintenance and supply cost segments.

A. Finding the Minimum Logistics System Cost Corresponding to Given Aircraft Costs

For a given flying activity program, the average relationship between the cost[9] of aircraft not operationally ready because of maintenance and the cost of maintenance support for keeping a specified number of aircraft operationally ready may look like curve A in Figure 1. The curve indicates that a desired number of operationally ready aircraft can be achieved by many combinations of maintenance support cost and cost of aircraft not operationally ready because of needed maintenance. The curve states that successive increments to expenditures on maintenance support will produce smaller and smaller reductions in the cost of aircraft not operationally ready because of needed maintenance. Also, because of the independence assumptions, this relationship is not affected by the amount of supply support.

---

[7] The "budget" as used here means the initial cost of aircraft plus operating costs for a given number of years, defined by the service life of the aircraft, plus capital costs in supply and maintenance and corresponding operating costs, also measured for the service life of the aircraft.

[8] The method to be described is somewhat more general than that required to solve the problem presented. For example, this more general method would have to be used if the assumption of independence between supply and maintenance were dropped or if the cost of aircraft depended on the number of aircraft purchased. Since further development of this subject will likely involve such changes in assumptions, it seemed preferable to present a more general procedure.

[9] The cost of aircraft not operationally ready as used here and elsewhere in this paper is the product of the number of such aircraft and the average cost of an aircraft. This average cost includes the average purchase price of an aircraft (including the effect of quantity produced) plus any additional costs that can be identified to the aircraft, such as special modifications.

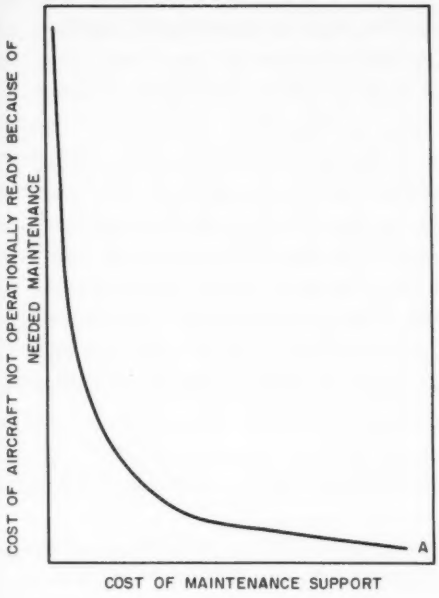


Figure 1 - Relationship between cost of maintenance support and cost of aircraft not operationally ready because of needed maintenance for attaining a specified number of operationally ready aircraft and a flying-hour program

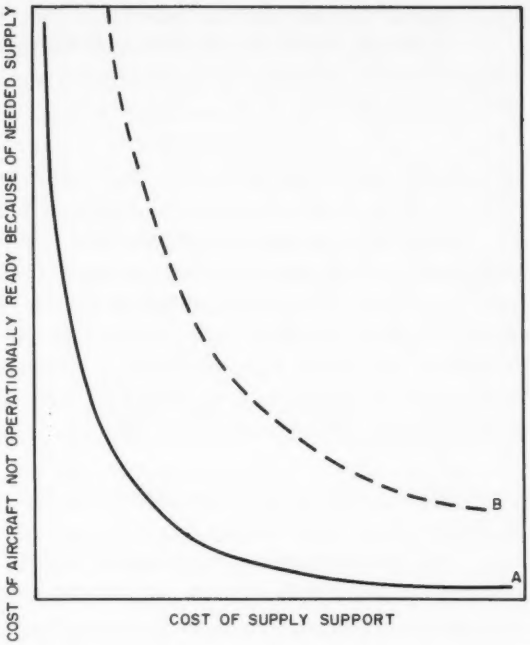


Figure 2 - Relationship between cost of supply support and cost of aircraft not operationally ready because of needed supply for attaining specified numbers of operationally ready aircraft and flying-hour programs

That latter relationship is illustrated in Figure 2. It has the same general appearance as the maintenance support relationship. Curve A provides the different combinations of supply support cost and aircraft not operationally ready because of needed parts that will meet the given objective of operationally ready aircraft and flying hours. We also indicate a curve B which shows the combination of supply support and aircraft not operationally ready because of needed parts that will meet a somewhat higher objective, measured in operationally ready aircraft and flying hours. This latter curve is given to indicate the effect on the curve's position if the desired objective is higher.

From curve A in Figures 1 and 2 we then compute another curve which gives the combinations of supply support cost and maintenance support cost that result in minimum logistics system cost (sum of supply support cost and maintenance support cost) for different sizes of aircraft inventory (measured in dollars of cost). The Appendix describes how the curve in Figure 3 that presents these combinations is computed. Each of the points on this curve thus gives the combinations of minimum logistics system cost (maintenance support costs plus supply cost) for given aircraft cost that satisfies the desired objective of operationally ready aircraft and flying-hour program. Note that the curve is asymptotic to the line LL^1. The line LL^1 is parallel to the horizontal axis, and it represents the cost of the given number of operationally ready aircraft. There cannot be fewer than this number of aircraft in the inventory, regardless of the amount of logistics support.

B. Finding the Combination of Logistics System Cost and Aircraft Cost that Minimizes Budget

The next step in the solution is to find the combination of aircraft cost and logistics system cost that minimizes the total budget (aircraft cost plus logistics system cost). The solution is graphically done in Figure 4. Curve A is the same curve as that shown in Figure 3. Line B is a straight line drawn at a 45° angle. Therefore, the abscissa is the logistics system cost and the ordinate of each point on this line also equals the logistics system cost. Now if we take any vertical line, such as RST passing through curves A and B, the distance $\overline{RS}$ measures the logistics system cost, and the distance $\overline{RT}$ measures the aircraft cost for that logistics system cost. If we now construct a curve C by adding to $\overline{RT}$ the length $\overline{RS}$, so that $\overline{TU} = \overline{RS}$ and $\overline{RU} = \overline{RT} + \overline{RS}$, then the length of $\overline{RU}$ constructed at any point R on the horizontal axis measures the total budget cost. We want that budget cost which is a minimum. This is given by point P, the lowest point on curve C, measured on the vertical axis. It consists of a logistics system cost given by the length $\overline{OM}$, which is the abscissa of point P and an aircraft cost given by length $\overline{MN}$, where $\overline{NP} = \overline{OM}$.

C. Breaking Down the Logistics System Cost of Minimum Budget Solution into Supply and Maintenance Support Cost

We next want to break the logistics system cost of the minimum budget solution into its components: supply support cost and maintenance support cost. We first subtract from total aircraft cost the cost of the number of operationally ready aircraft. The residual represents the cost of aircraft not operationally ready. With this value, we can then manipulate the curves shown in Figures 1 and 2 in accordance with the method described in the Appendix, so that for that cost of aircraft not operationally ready, we obtain the corresponding breakdown of logistics system cost into maintenance support cost and supply support cost. This gives the results desired.

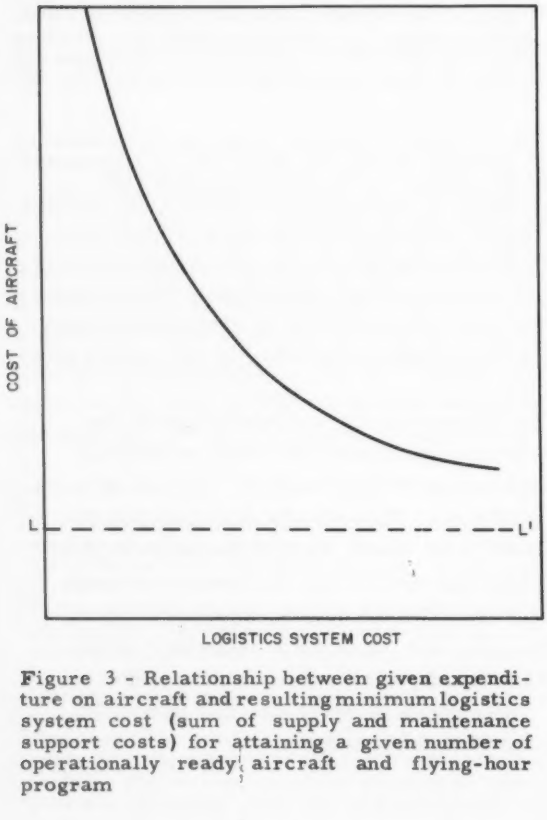


Figure 3 - Relationship between given expenditure on aircraft and resulting minimum logistics system cost (sum of supply and maintenance support costs) for attaining a given number of operationally ready aircraft and flying-hour program

D. Improved Allocations

The solution of the optimal allocation problem requires that we know the parameters of the curves used. This means that we must be able to determine the outcome of a great many different allocations of resources between aircraft, supply support and maintenance support. In practice, this is a very complicated problem because of lack of data or because the interaction among these several elements is very difficult to calculate. It may require extrapolations into situations for which there is no experience.

It may be much easier to find improved allocations (compared to a present, actual one) than the optimal allocation itself; and the establishment of such improved allocations could be a real step forward. The starting point for finding such allocations is the present allocation. We may determine whether, within the present budget, shifting resources from the logistics system to aircraft or from aircraft to the logistics system would produce a larger number of operationally ready aircraft. We could do this by investigating how many more operationally ready aircraft the present system would produce if an additional number of aircraft were added to the inventory and the logistics system budget were reduced by the cost of these aircraft, thus keeping the total budget constant. We could then compare the number of additional operationally ready aircraft that would be produced if the money equivalent to these additional aircraft were spent on the logistics system and the number of aircraft in the inventory decreased by that number. The shift producing the larger number of operationally ready aircraft would represent the improved allocation. This process may be repeated by increasing the number of aircraft shifted in one direction or the other until a reversal in the advantage is obtained. This would be the optimal allocation. For the reasons mentioned earlier, however, it may be possible to consider only limited shifts, and these may fall short of the optimum.

In the same way, shifts in allocations among aircraft, maintenance support, and supply support can be studied to find improved allocations of budget among these three elements. The purpose of this type of analysis would be to determine, for example, whether a combat unit that now contains 25 aircraft and a specified number of maintenance personnel would be able to attain a higher number of aircraft operationally ready if it consisted of 20 aircraft, with the money equivalent of the other five aircraft going into the hire of more maintenance support personnel.

V. CONCLUSIONS AND LIMITATIONS

This study describes some of the concepts involved in relating weapons to logistics expenditures. It presents principles of finding optimum or, at least, improved allocations of budget funds between maintenance support and supply support (as defined) and between the sum of the two and additional weapons. The ideas are both fundamental and extremely simple. Clearly the concept of balancing the inventory of a weapon against its support need should be considered at the time of planning its introduction and throughout its life. The paper presents only an approach to the problems, however, not a final solution. Some major limitations of the method described must be kept in mind, besides certain assumptions made in the text.

First, the objective of the logistics system is defined as maximizing the

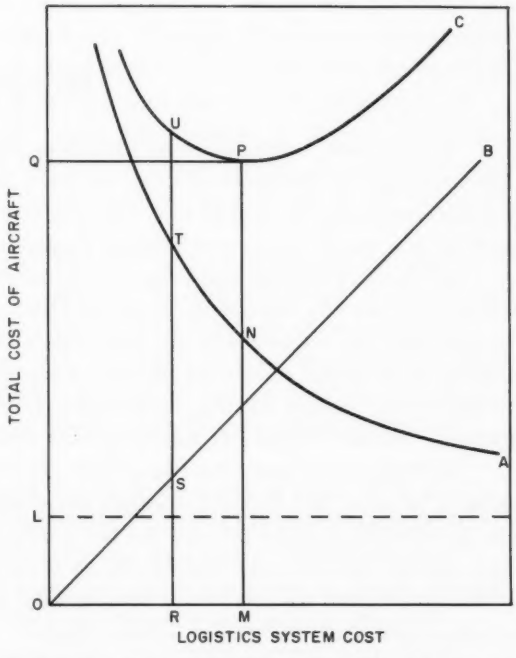


Figure 4 - Combination of logistics system cost and aircraft cost that minimizes the total budget

number of operationally ready aircraft, given the flying program. In effect, this implies that aircraft not operationally ready have no value. In a situation in which they could be made operational, nonoperational aircraft would, of course, have some value. Similarly, the value of the operational readiness of a weapon may vary somewhat between commands. Moreover, during the life cycle of an aircraft this value may change considerably. These and other dynamic elements in the life of a weapon are disregarded in this study.

The technique outlined in Section IV is not useable at this time because of the lack of suitable data. We can surmise that the curves have the general characteristics described, but we cannot answer the real question of the extent to which nonoperationally ready levels will be changed by shifts in resource use. It is particularly difficult, in fact impossible, to get reasonable approximations over any large segments of the curves from existing data. With additional research and improved data systems, it may be possible to define short stretches or a few points on the curves for major weapons.

Two other things should be mentioned. Not all resource limitations are reflected in dollars. For example, with personnel ceilings it may be necessary to incur higher dollar costs rather than higher manpower costs in supporting a weapon. Lastly, the relationships between different weapons systems are ignored. Different weapons do make joint use of many facilities and stocks of parts and equipment. This is an important problem, but the approach outlined here does not take account of the interplay.

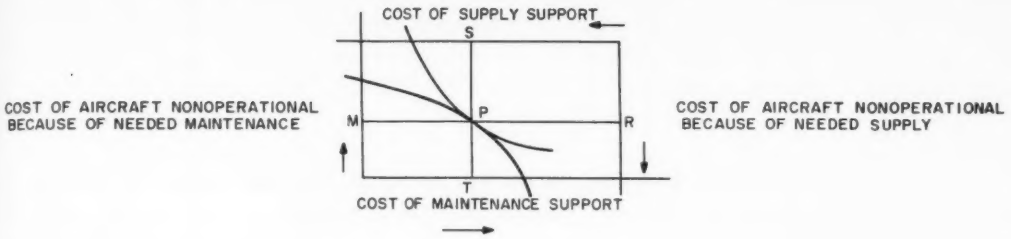
An adequate understanding of the concepts may be helpful in itself, and for this reason the present study is being published. In the meantime, efforts are continuing at RAND to develop quantitative results that could be of more direct benefit to policy-making.

APPENDIX

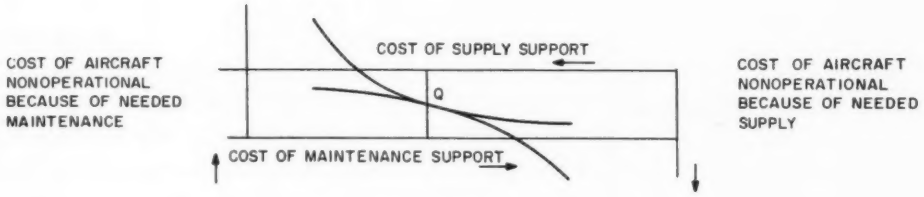
The purpose of this Appendix is to describe how the logistics system relationships used in the main body of this paper are derived from the separate relationships for supply support and maintenance support and then to describe how the optimal allocation for the logistics system as a whole is subdivided between supply support and maintenance support.

We assume that the maintenance support and supply support relationships are available and that they look like the curves shown in Figures 1 and 2. The principle of optimal allocation is that we allocate resources to the two forms of support so that the spending of an additional dollar on either would reduce the number of aircraft not operationally ready because of needed maintenance (or supply) equally. Consequently, whenever the supply support and maintenance support curves are tangent, we get an optimal allocation.

In Figure 5, we show the two curves involved in two positions of tangency. To secure tangency, we have first turned the supply relationship curve upside down; second, we have fixed the vertical distance between the two horizontal axes. This, in effect, fixes the total cost of aircraft not operationally ready (because of either needed supply or needed maintenance). We have then moved the two curves horizontally, maintaining the vertical separation of the horizontal axes, until the curves become tangent. Point P in Figure 5 is one such point. Point Q is another, found by using a smaller vertical distance. Each of these points is characterized by the condition that for the given amount of nonoperational aircraft, as



(a) First illustrative calculation


COST OF AIRCRAFT NONOPERATIONAL BECAUSE OF NEEDED MAINTENANCE
COST OF SUPPLY SUPPORT
Q
COST OF AIRCRAFT NONOPERATIONAL BECAUSE OF NEEDED SUPPLY
COST OF MAINTENANCE SUPPORT

(b) Second illustrative calculation

Figure 5 - Illustrative calculations of total logistics system cost for given cost of nonoperational aircraft inventory

measured by the vertical separation of the horizontal axes, they identify the minimum cost of logistics support required to meet the specified objective. This cost of logistics system support is geometrically given by the horizontal separation of the vertical axes with, for example, in the case of Point P, $\overline{MP}$ measuring the cost of maintenance support and $\overline{PR}$ the cost of supply support. In the same way, the distance $\overline{PS}$ measure the number of aircraft not operationally ready because of needed supply and $\overline{PT}$ the number of aircraft not operationally ready because of needed maintenance. Similar interpretations apply to point Q for its set of conditions.

Now, we should like to describe how Figure 3 is derived from the points obtained as shown in Figure 5. The vertical axis in Figure 3 measures the cost of aircraft in the system. This cost is given by the sum of the cost of operational aircraft (which is specified by the objective) plus the cost of nonoperational aircraft which is obtained from solutions of the type described in Figure 5. The horizontal axis in Figure 3 is the cost of logistics support, and these values are also obtained from figures like Figure 5, with the corresponding cost of nonoperational aircraft.

Thus, in Figure 6, which is a duplication of Figure 3, the curve is obtained by adding to each cost of nonoperational aircraft derived from Figure 5 the fixed cost of operational aircraft, leaving the corresponding cost of logistics support unchanged. This addition is indicated on the vertical scale of Figure 6. We have also plotted the P and Q points derived in Figure 5 to show the connection between Figure 6 and Figure 5 and, in turn, Figure 3.

Now we shall discuss how to obtain the breakdown of the total aircraft cost, that is, part of the minimum budget solution derived by Figure 4, into its components: cost of

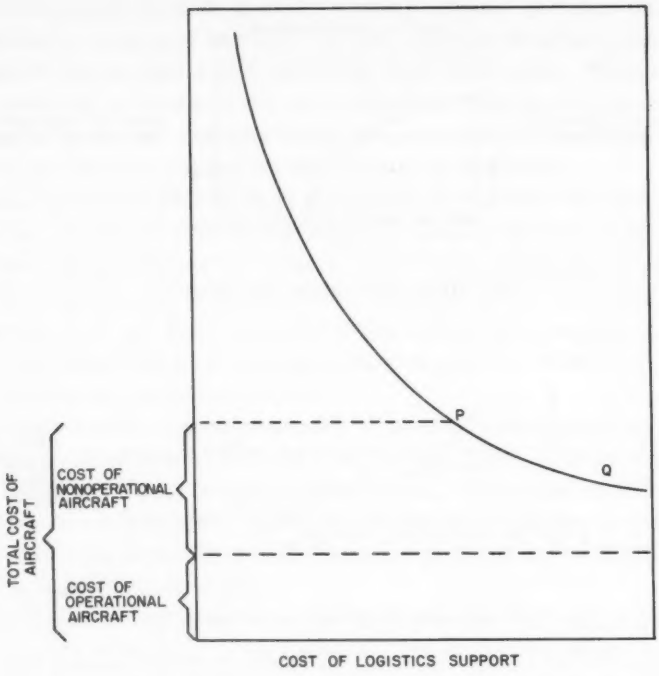


Figure 6 - Derivation of total logistics system relationships curve from maintenance and supply relationships

operational aircraft, cost of aircraft nonoperational because of needed maintenance, and cost of aircraft nonoperational because of needed supply. From Figure 4, the solution of the minimum budget with its components of cost of aircraft and cost of logistics support is obtained. Now, returning to Figure 6, the cost of nonoperational aircraft is obtained by subtracting the cost of operational aircraft from total aircraft cost. Next, to break down the cost of nonoperational aircraft into its components, we use the technique illustrated in Figure 5. We place the two curves used in Figure 5 (in either illustrative calculation) so that their horizontal axes are separated vertically by the cost of nonoperational aircraft just obtained and the vertical axes are separated by the corresponding cost of logistics support. Under these two conditions, the two curves should be tangent because of the relationship between Figures 4 and 5. Also, the vertical distance between the point of tangency and the lower horizontal axes gives the cost of aircraft nonoperational because of needed maintenance, and the vertical distance between the point of tangency and the upper horizontal axes gives the cost of aircraft nonoperational because of needed supply, just as occurs in Figure 5.

* * *

AN APPLICATION OF GAME THEORY TO SPECIAL WEAPONS EVALUATION

J. K. Hale and H. H. Wicke*
Sandia Corp., Albuquerque, N.M.

INTRODUCTION

Current (1956) published literature pertaining to special weapons selection and evaluation makes it appropriate to preface this paper with a few remarks on the use of a game-theoretic approach and on the motivation for the use of Lanchester theory. One currently much-employed approach to weapons evaluation is to choose models for weapon and target and to calculate damage. The results of a careful calculation, judiciously handled, can be of assistance in weapons evaluation. It is reasonable to assume, however, that in future wars both sides will have atomic weapons and both sides will be forced to develop protective measures to survive. A major means of protection which has been considered here to the exclusion of others is the manner in which opposing troops are disposed on the battle field.

That troop disposition is important can be seen from simple considerations. If enemy troops are in a long, thin line close to friendly troops, a large-yield weapon employed by friendlies will be ineffective, while several small-yield weapons can be quite effective. On the other hand, if troops are more widely dispersed, a large yield will be more effective than a small yield. (The military effect of the troop dispersion will be discussed in a later paragraph.)

These considerations lead to questions as to yields and troop formations. In this paper a simple battle situation in which these elements enter has been devised in order to get some rough indications of what answers to expect to these questions. The present model can be refined in many ways, but even this first approximation points out some trends and shows need for further thinking on the subject.

The most important point is, however, that one needs answers to these questions (as well as many others) to avoid disaster. As can be readily shown, the use of one kind of weapons system or of one kind of troop formation can lead to defeat. One needs only to reflect that, given a weapon system, one can usually develop tactics which will render that system ineffective (unless one is forced by other considerations, such as terrain, to maintain a certain formation). Thus it appears that at least two different weapons systems and two different troop formations should be employed. To find out exactly what strategies to use, the ideal procedure would seem to be to set up a continuous game in which each player's strategy

*Manuscript received February 19, 1957

ranges over all possible combinations of weapons and troop dispositions, subject to certain restrictions, e.g., budget, number of troops, terrain, military objective, etc. This game is, of course, far beyond present computational capabilities, and a very much simplified version is treated here.

Some restriction must be put on the formations that can be employed. For example, a good defense against an atomic weapon is to disperse troops by placing men so that any given man is 100 miles away from the nearest man of the remaining troops. This is clearly militarily ineffective, however, and Lanchester theory is used to judge the effectiveness of a military formation. The body of this paper treats the way in which the theory is employed to handle formations.

METHOD

Assume that each man in a force has a certain fire power independent of the size of the group. Further, a "basic" group exists for both the attacker and defender in the sense that it is the most desirable operational size of a group. The number of groups in a certain formation of troops is thus measured in terms of multiples of this fundamental group and is a function of the total number of men available and the number of men required for a fundamental group. Consequently, after a weapon system has been employed, one can then assess the damage to each of these basic groups. Suppose that the attacking force has n basic groups and, after a weapon system has been employed by the defending force, each attacking group A_i contains n_i men, each with firepower K.[1] Further, suppose the defending force has m basic groups and, after a weapon system has been employed by the attacking force, each defensive group D_i consists of m_i men, each with fire power L. The remaining groups A_i and D_i are then allowed to make engagements (the method of engagement will depend upon the location of the groups) one at a time, and the expected number of men destroyed, as a function of the time of the engagement, satisfies the Lanchester decay law; i.e., if a group A_i engages group D_i, and the expected numbers of men at any time, t, of the attacking and defending forces are y and x, respectively, then

$$\frac{dy}{dt} = -Lx$$

and (1)

$$\frac{dx}{dt} = -Ky ,$$

with the initial conditions

$$y(0) = m_i$$

and

$$x(0) = n_i .$$

By taking into account the time at which groups can engage, regroup, etc., and by a repeated application of this law, one can determine who wins; or, more precisely, one can determine

[1] One could allow the firepower per man to change with the number of men in the group, but we assume in this paper that this firepower is constant.

the expected number of men remaining on each side as a function of time. If one defines "annihilation time" as that time at which one of the forces looses all its men, one could consider the payoff from such an encounter as the difference between the number of men remaining for the attacking force and the number remaining for the defending force at the annihilation time.

If the expected number of men remaining is not a desirable payoff function, one could use the more general Lanchester equations for the probability that a certain number of men remain for the attacking and defending forces at time t. For a discussion of these more general equations, see J. R. Isbell and W. H. Marlow, "Attrition Games," Naval Research Logistics Quarterly, Vol. 3, 1956, pp 71-93. In this paper, to illustrate the ideas, only Equations (1) will be considered.

The calculations involved in the general procedure outlined above would be extremely complicated and, therefore, it is assumed in the remainder of this paper that the number of men remaining immediately after the engagement of two groups is determined by the solution of Lanchester's Equations (1) at the annihilation time. Also, it is assumed that the errors in the delivery of the weapons are zero, and the ratio, K/L, of the firepowers is one.

The remainder of the paper is devoted to a discussion of three examples which illustrate how this method may be applied. Although the division into basic groups is artificial, it is felt that the results are still informative.

EXAMPLES

1. Atomic Weapons for Defensive Forces Only

Consider the close-support situation in which only the defensive forces have access to atomic weapons. The defensive forces are supposed to consist of M men, each with firepower K, and these men are in a rectangle 600 x 3600 yards. In this problem, the configuration of the defensive forces is not important. They are assumed to be in a rectangle so that this same model can be used in the later problems. The attacker, with a total of N men, each with firepower L, is assumed to assemble his men into basic rectangular groups of dimension 600 x 3600 yards, but he is allowed to have many such groups. Furthermore, the attacker always assembles his basic groups in the manner shown in Figure 1. The firepower per man, L, of each man in the attacking force is assumed to be independent of the dispersion of the total number of troops, N. Furthermore L/K = 1.

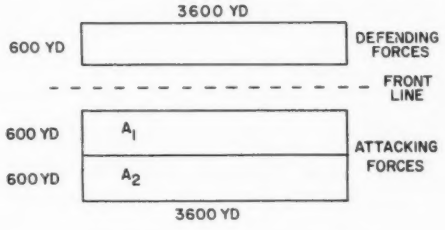


Figure 1 - Battle situation when the attacker has two groups, A_1 and A_2

A strategy for the defender is the choice of a weapon system, whereas the strategy for the attacker is the choice of the number, W, of basic groups in which to divide his men. For simplicity in the calculations, the defender must choose among the four weapon systems of Table 1, and the attacker can divide his forces into only one, two, three, or four groups. Furthermore, each weapon system has a fixed set of aim points, chosen on a line parallel to the front line (Figure 1) and chosen so that the friendly troops will be safe. Assuming zero delivery errors, one can then easily calculate the effectiveness of a given weapon system against any of the formations the attacker may choose to employ; i.e., if the formation consists of k basic groups, one can calculate the number of men, N_{ij}, remaining in the group A_i,

$i = 1, 2, \ldots, k$ of the attacker after the defender has employed a weapon system j, $j = 1, 2, \ldots, 4$. The numbers N_{ij} are given in Table 2.

TABLE 1

Weapon Systems Available to the Defender

Weapon system	1	2	3	4
Number of weapons	1	2	3	4
Effect radius of each weapon (yds)	1100	950	850	550

TABLE 2

Number of Men Remaining in Each Group A_i of Attacker's Force After Employment of Weapon System by Defender When Attacker Has k Basic Groups and Total of N Men

No. of Group, k	Group	Weapon System			
		1	2	3	4
1	A_1	0.67N	0.3N	0.348N	0.061N
2	A_1	0.33N	0.153N	0.174N	0.031N
	A_2	0.19N	0	0	0.197N
3	A_1	0.22N	0.091N	0.116N	0.021N
	A_2	0.126N	0	0	0.129N
	A_3	0.126N	0.031N	0.087N	0.33N
4	A_1	0.167N	0.076N	0.087N	0.015N
	A_2	0.095N	0	0	0.098N
	A_3	0.095N	0.035N	0.087N	0.25N
	A_4	0.167N	0.227N	0.25N	0.25N

The Lanchester theory is applied in the following manner. Suppose the attacker has chosen two basic groups and the defender has chosen weapon system 1. After the defender employs the weapon system, the attacker has 0.33N and 0.19N men remaining in Groups A_1 and A_2, respectively. It is then assumed that Group A_1 is the first to attack the defender, then A_2, and each new attack is made against the remaining men of the defender. For any such attack, the Lanchester theory always leads to complete annihilation of one group and the number of men remaining after all attacks have been made are recorded in the appropriate place in Table 3. The negative numbers in the table correspond to annihilation of the attacking force; the function[2] $g(x) = (\operatorname{sgn} x)\sqrt{|x|}$, and $H = (M/N)^2$. (The numbers in the table have been normalized on N, but this does not affect the strategies in the game.)

[2]The term sgn x is equal to the algebraic sign of x.

TABLE 3

General Payoff Matrix for the Game in Which Only the Defender Has Atomic Weapons, $g(x) = (\text{sgn } x)\sqrt{|x|}$, $H = (M/N)^2$

No. of Groups, k	Weapon System			
	1	2	3	4
1	g(0.445 - H)	g(0.090 - H)	g(0.121 - H)	g(0.004 - H)
2	g(0.145 - H)	g(0.023 - H)	g(0.030 - H)	g(0.040 - H)
3	g(0.070 - H)	g(0.009 - H)	g(0.021 - H)	g(0.127 - H)
4	g(0.074 - H)	g(0.058 - H)	g(0.078 - H)	g(0.135 - H)

Observe that if $x_2 > x_1$, then $g(x_2) > g(x_1)$. For if $x_2 > x_1 > 0$, then $g(x_2) - g(x_1) = \sqrt{x_2} - \sqrt{x_1} > 0$; if $x_2 > 0 > x_1$, then $g(x_2) - g(x_1) = \sqrt{x_2} - \sqrt{|x_1|} > 0$; and if $0 > x_2 > x_1$, then $|x_1| > |x_2|$ and $g(x_2) - g(x_1) = -\sqrt{|x_2|} + \sqrt{|x_1|} > 0$. From this fact, it is easily seen that for every value of H the element in the second column and the j^{th} row of the matrix of Table 3 is less than each of the elements in the first and third columns of the j^{th} row for j = 1, 2, 3, and 4. Therefore, the defender will never use weapon system 1 or 3, and the above game is equivalent to the game with payoff matrix given in Table 4.

Using the same reasoning as above, one sees that the element in the fourth row and the j^{th} column of the matrix in Table 4 is greater for every value of H than each of the elements in the second and third rows of the j^{th} column for j = 1 and 2. Therefore, the attacker will never divide his troops into either two or three groups, and the game is finally reduced to the game with payoff matrix given in Table 5.

TABLE 4

General Payoff Matrix for the Game Equivalent to the Game in Which Only the Defender has Atomic Weapons, $g(x) = (\text{sgn } x)\sqrt{|x|}$, $H = (M/N)^2$

No. of Groups, k	Weapon System	
	2	4
1	g(0.090 - H)	g(0.004 - H)
2	g(0.023 - H)	g(0.040 - H)
3	g(0.009 - H)	g(0.127 - H)
4	g(0.058 - H)	g(0.135 - H)

TABLE 5

General Payoff Matrix for the Game Equivalent to the Game in Which Only the Defender has Atomic Weapons, $g(x) = (\text{sgn } x)\sqrt{|x|}$, $H = (M/N)^2$

No. of Groups, k	Weapon System	
	2	4
1	g(0.090 - H)	g(0.004 - H)
4	g(0.058 - H)	g(0.135 - H)

For any value of H, an optimal strategy for the attacker consists of specifying the probabilities of dividing his troops into four groups and of not dividing his troops, whereas an optimal strategy for the defender is the probabilities of using weapon systems 2 and 4.

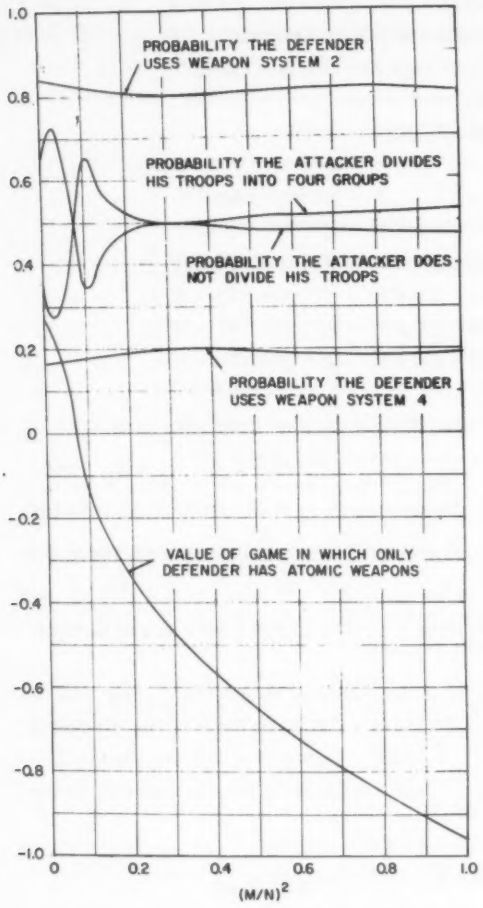


Figure 2 - Solution of the game in which only the defender has atomic weapons, as a function of $(M/N)^2$

Figure 2 shows these optimal strategies, together with the value of the game as a function of $(M/N)^2$. Note that the solution of the general 4×4 game in Table 3 does not involve strategies 2 and 3 for the attacker and strategies 1 and 3 for the defender.

A few points on this curve are of particular interest. For example, the point at which the value of the game is zero may be taken as a measure of the value of an atomic weapons system in terms of men. More specifically, if the defender has access to atomic weapons and the attacker does not, then the defender needs only approximately 0.28 as many men as the attacker to have the same capability.

Also, if the attacker and defender initially have the same number of men, say 1000, then the expected number of men remaining for the defender is 950 and for the attacker is zero.

2. Atomic Weapons for Both Offensive and Defensive Forces

In this section, the battle situation will be the same as discussed in Example 1, except that the defending forces will now be allowed to divide into groups. Further, it is assumed that the defensive forces will be better protected against atomic weapons and, therefore, the radii of effect of the attacker's weapons are taken to be smaller than those for the defender. The available weapon systems for the attacker and defender are given in Table 6. Both the attacker and defender are allowed to present their forces in either one or four basic groups. (A basic group is a rectangle 600×3600 yards.)

If one applies the method described in Example 1 and lets (i, j) denote a strategy which consists of a choice of weapon system j and a formation composed of i groups, then one can calculate the payoff matrix of Table 7.

From the reasoning in Example 1, it is easily seen that for every value of H the defender always prefers strategy (1, 1) to (4, 1), and strategy (1, 2) to (4, 2). Therefore, this 4×4 rectangular game is equivalent to a 4×2 game, with the strategies for the defender being (1, 1) and (1, 2). Furthermore, in this reduced game, the attacker always prefers (1, 2) to (1, 1), and (4, 2) to (4, 1). Consequently, the game is reduced to an equivalent game with payoff matrix given by Table 8.

TABLE 6
Weapon Systems Available to Attacker and Defender

	Attacker's weapon systems		Defender's weapon systems	
	1	2	1	2
Number of weapons	2	4	2	4
Effect radius of each weapon (yds)	720	400	950	550

TABLE 7
General Payoff Matrix for the Game in Which Both Defender and Attacker Have Atomic Weapons, $g(x) = (\operatorname{sgn} x)\sqrt{|x|}$, $H = (M/N)^2$

Strategies for the Attacker	Strategies for the Defender			
	(1, 1)	(1, 2)	(4, 1)	(4, 2)
(1, 1)	g(0.090 - 0.620H)	g(0.004 - 0.620H)	g(0.090 - 0.101H)	g(0.004 - 0.101H)
(1, 2)	g(0.090 - 0.204H)	g(0.004 - 0.204H)	g(0.090 - 0.163H)	g(0.004 - 0.163H)
(4, 1)	g(0.056 - 0.620H)	g(0.135 - 0.620H)	g(0.056 - 0.101H)	g(0.135 - 0.101H)
(4, 2)	g(0.056 - 0.204H)	g(0.135 - 0.204H)	g(0.056 - 0.163H)	g(0.135 - 0.163H)

TABLE 8
General Payoff Matrix for the Game Equivalent to the Game in Which Both Defender and Attacker Have Atomic Weapons, $g(x) = (\operatorname{sgn} x)\sqrt{|x|}$, $H = (M/N)^2$

Strategies for the Attacker	Strategies for the Defender	
	(1, 1)	(1, 2)
(1, 2)	g(0.090 - 0.204H)	g(0.004 - 0.204H)
(4, 2)	g(0.056 - 0.204H)	g(0.135 - 0.204H)

Table 8 shows that for any value of H an optimal strategy for the attacker consists of the employment of weapon system 2 and the specification of the probabilities of dividing his troops into one and four groups, whereas the defender never divides his troops and must specify the probabilities of employing weapon systems 1 and 2. Figure 3 shows these optimal strategies together with the value of the game as a function of $(M/N)^2$. Note that the solution of the general 4×4 game in Table 7 does not involve strategies (1, 1) and (4, 1) for the attacker and strategies (4, 1) and (4, 2) for the defender.

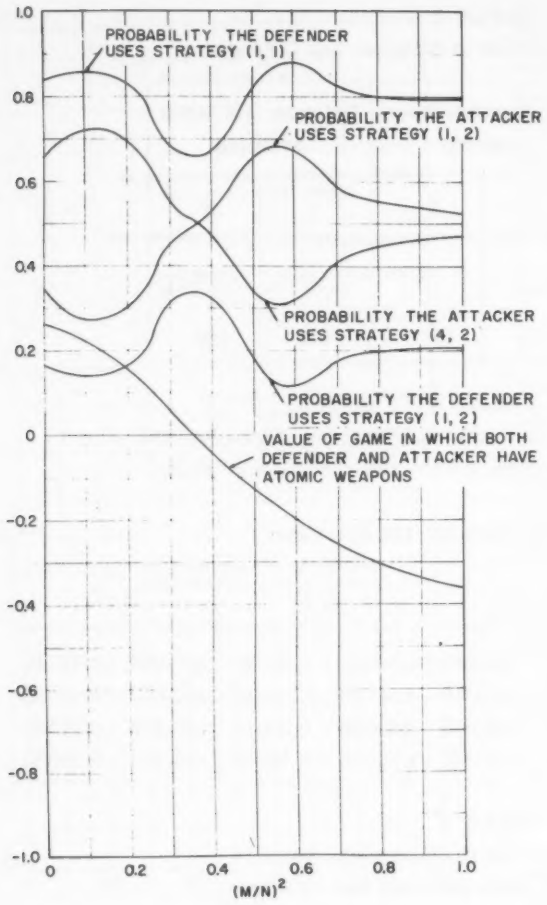


Figure 3 - Solution of the game in which both the defender and the attacker have atomic weapons, as a function of $(M/N)^2$

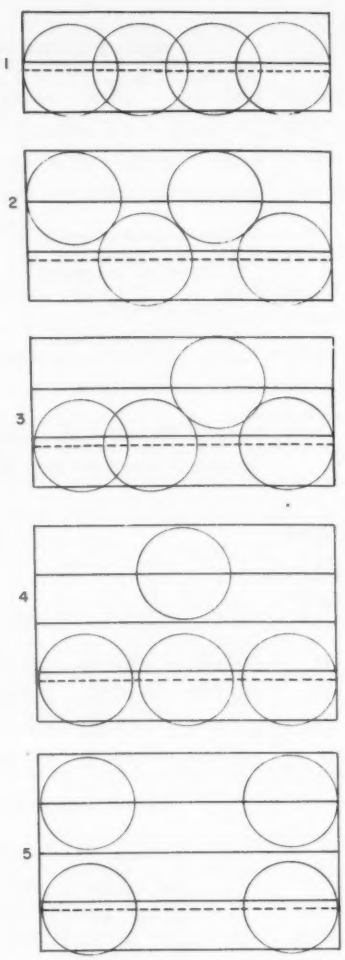


Figure 4 - Aiming procedures for the defender. Rectangles represent "basic" groups of attacker's forces. Circles represent effects of weapons used by defender. Each number indicates a different situation. Defender's forces (not shown) are below attacker's forces.

The point on Figure 3 at which the value of the game is zero may be taken as a measure of the value of being prepared for an atomic burst in terms of men. More specifically, if both the defending and attacking forces have access to atomic weapons and the defending forces are unwarned, then the defending forces need approximately 0.60 as many man as the attacker to have the same capability.

Also, if the attacker and defender initially have the same number of men, say 1000, the expected number of men remaining for the defender is 360 and for the attacker is zero.

3. Optimal Aiming Procedure When Only the Defender Has an Atomic Weapon System

In Example 3, the battle situation is the same as that discussed in Example 1, except the defender will have only one weapon system consisting of four weapons each with an effect radius of 550 yards. The defender will have a choice of five aiming procedures, as shown in Figure 4.

By applying the methods described in Example 1, one can show that the payoff matrix is given by Table 9.

TABLE 9

General Payoff Matrix for the Game in Which the Defender Has an Atomic Weapon System and Various Aiming Procedures, $g(x) = (\text{sgn } x)\sqrt{|x|}$, $H = (M/N)^2$

No. of Groups, k	Aiming Procedure				
	1	2	3	4	5
1	g(0.445 - H)	g(0.226 - H)	g(0.090 - H)	g(0.075 - H)	g(0.265 - H)
2	g(0.145 - H)	g(0.078 - H)	g(0.036 - H)	g(0.060 - H)	g(0.158 - H)
3	g(0.070 - H)	g(0.065 - H)	g(0.080 - H)	g(0.096 - H)	g(0.100 - H)
4	g(0.074 - H)	g(0.098 - H)	g(0.108 - H)	g(0.093 - H)	g(0.079 - H)

It is clear that the attacker always prefers having his troops in one group to having his troops in two groups. Consequently, the game is equivalent to the 3×5 game obtained from Table 9 with the second row deleted.

This game is solved for only one value of H, namely, $H = 0.08$. The value of the game is 0.075, and the attacker uses strategies 1, 3, and 4 with probabilities 0.24, 0.45, and 0.31, respectively, where the defender uses strategies 1, 3, and 4 with probabilities 0.21, 0.03, and 0.76, respectively.

It is interesting to note the variation of the value of this game to certain reduced games for $H = 0.08$. For example, if the defender is allowed only strategies 1 and 4, then the value of the game is 0.077. If he is allowed only strategies 1, 2, and 4, then the value of the game is 0.076.

Notice also that the optimal aiming procedure does not involve only the aiming procedures which are optimal against each of the fixed dispersion patterns for the attacker. In this case, suboptimization is not valid.

CONCLUSIONS

A method was introduced for combining the techniques of classical Lanchester theory of combat with those of game theory toward the end of selecting optimal strategies in combat with special weapons. In the application of this method to the example in which only the defender had atomic weapons, it was shown that the attacker always chose either to disperse his troops the maximum amount or not to disperse his troops at all. The defender always chose to employ a mixed strategy consisting of the weapon systems of either two intermediate weapons or four small weapons.

If both the defender and attacker had access to atomic weapons, then the optimal strategy for the attacker was to employ the weapon system consisting of four small weapons and to use a mixed strategy for the dispersion of his troops. On the other hand, the defender never dispersed his troops and always used a mixed strategy for the weapon systems.

In the example where the defender has a fixed weapon system and chooses to optimize his aiming procedure, it was shown that the optimal aiming procedure does not involve only the aiming procedures which are optimal against each of the fixed dispersion patterns for the attacker.

The model discussed in this paper is far from realistic, but the authors feel that certain interesting trends may be obtained by such elementary discussions. Two ways in which to approach more realism are to introduce into the combat the time at which the different groups become engaged, as discussed at the beginning of the section on "Method," and to obtain a more realistic model for the basic group discussed within the same section.

* * *

AN OPTIMAL SEARCH PATTERN[1]

J. R. Isbell*
The Institute for Advanced Study

The problem of minimax-time search in a plane for a line at known distance is solved by a path made up of line segments and circular arcs. Bellman posed this and some related problems [1]; a parallel strip problem is partially analyzed below. It may be noted that no systematic method for such problems is available. A solution is typically piecewise smooth, and the main difficulty is in determining how many pieces there are and what roles they play.

1. ONE LINE

Given a base point p and a line L at unit distance from p in a plane, to find a rectifiable curve C of minimum length such that C joins p to L, and if C is rotated about p through any angle, it still meets L.

In other words, one is at known distance from the line L but does not know its direction. It is required to find the shortest path which will surely get to L. It is equivalent to require that the path C meet every tangent to the unit circle about p or that the convex hull of C contain the circle.

Note that a radius and circle of total length $2\pi + 1 \sim 7.283$ will find L, or indeed any set at unit distance from p. Among all convex sets it can be shown that a line is the easiest to find, and this search turns out to require $7\pi/6 + \sqrt{3} + 1 \sim 6.397$ units.

Being at p and unoriented, imagine a clock face. Walk toward one o'clock for $\sqrt{4/3}$ units. (This takes you to a vertex of a circumscribed regular hexagon.) Then turn on the tangent which strikes the unit circle at two o'clock. Follow the circle to nine o'clock and continue on a tangent. Upon striking the line which is tangent to the unit circle at twelve o'clock you have swept all tangents to the unit circle, and that in a path of minimum length.

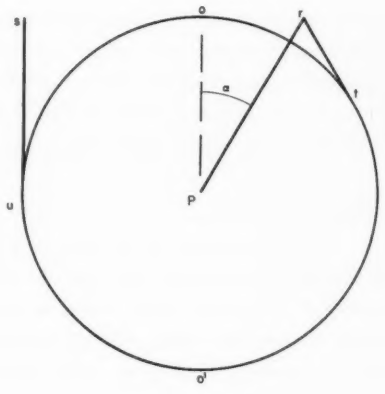


Figure 1

Proof in outline, supposing a smooth solution exists. Define the rim of the solution curve C to consist

*Manuscript received December 9, 1956
[1]Research supported by the Office of Naval Research under Contract N7onr-41904

of all points q of C such that q is maximally distant from p on the ray pq. Clearly every rotation of the rim meets L. Let r be the first point of the path C in the closure of the rim. Then C begins with the line segment pr. We defer the question as to whether or not the path then includes a stretch of the rim; suppose it does. Then C never leaves the rim, as is seen by a local application of the triangle inequality; for if C leaves the rim we have the points r, r' not in the rim and r'' beyond r' on the ray pr', traversed by a path at least as long as two sides of the triangle. Therefore C consists of the segment pr and the rim rs.

Again by a local argument, rs is a convex curve. Its convex hull contains the unit disc about p, and hence if rs is a closed curve (s = r) then it has length at least 2π. Since pr has length at least 1, this implies that the length of C is greater than the length of the curve already described above, which is absurd. Thus r and s are different points, both outside the unit circle. The rim rs touches the circle, as otherwise we could decrease its length. If t is the first point after r where rs touches the circle, then by a similar argument t is no farther than the foot of the tangent from r. Since the rim is convex, rt is the tangent. Similarly C goes from the circle to s by a tangent. By a local argument, between the feet of these tangents C coincides with the circle. Thus C consists of the segment pr, the tangent segment rt, an arc of circle tu, and the tangent us. Finally, the line segment sr is another tangent to the circle. Therefore C is determined by two angles, the angular positions of r and s relative to the point o at which sr meets the circle.

These two positions are found by minimizing an elementary function of two variables, but it saves some work to note that the antipodal point o' to o on the circle is necessarily in C. Therefore the parts of C on the two sides of the diameter oo' may be treated separately. The length pr + rt + to' is given by $\sec \alpha + \tan \alpha + \pi - 2\alpha$, where 2α is the angle opt; this is a minimum at $\alpha = \pi/6$. Similarly, the two tangents from s subtend a right angle. Therefore the solution is the curve previously described.

To convert this sketch into a rigorous proof is a routine exercise. Instead of supposing a solution exists, one applies the argument to any other suitable curve C' to transform it into C by reducing its length. To get an interval of the path in its rim one observes that locally the path can be shortened by being made non-strictly convex away from p, if it is not already so. A convex function is piecewise differentiable, and then routine (but somewhat lengthy) arguments establish that the rim contains an interval. From this interval one can proceed, shortening C' steadily until it assumes the form pr + rs, where pr is a segment and rs the rim of the path. The rest of the argument offers no difficulty.

2. TWO LINES

If one is given L, as above, and a line L' parallel to L at unit distance from p in the opposite direction, then one can find one of these lines in about 3.256 units. Even if one must search for one of the lines and then go to the other one, the length (5.256) is considerably shorter than 6.397. But this is to begin with considerable information. Bellman [1] has asked for the minimax path, given only that p is somewhere in the strip of known width 2 between L and L'.

The only contribution here is the remark that the solution of this problem is likely to be an envelope of the family of solutions of the somewhat easier problems, to find L or L', given that the distance to L is 1 - t and the distance to L' is 1 + t. I am unable, however, to solve these problems. Clearly, for 1 - t sufficiently small it will be optimal to search for L alone by the search pattern given above. The solution for the opposite extreme t = 0 is given below.

It seems likely that the solutions for all t will split into cases; as remarked above, there is no systematic method for finding the critical places where such a problem splits, and the author's ingenuity is inadequate.

It should be noted that in [1] Bellman also asks for some search patterns minimizing expected time rather than maximum time.

We describe the solution of the case $t = 0$ of the above problem; the proof may safely be left to the interested reader. Walk toward one o'clock for $\sqrt{4/3}$ units and then follow the tangent to two o'clock (as before). Follow the circle to three o'clock and then leave on the tangent; either L or L' must appear before you reach the line tangent to the unit circle at six o'clock.

REFERENCE

[1] R. Bellman, Minimization Problem, Bull. Amer. Math. Soc. 62 (1956), p. 270

* * *

NOTES

The following has been received:

October 25, 1957

Dear Sir:

I am sorry to report an error in my recent paper published in the Quarterly, v. 4, n. 1, pp. 79 - 85. The sentence immediately after inequation (5.5), p. 83 is incorrect. Inequality (5.7) on the same page and assertions (A5) and (A6) on p. 84 are based on this statement and are also incorrect. Recognition of the error makes most of the subsequent discussion irrelevant and leaves (A4), p. 82 in doubt.

Mr. Donato A. D'Esopo of the Stanford Research Institute noticed the error and provided a counter-example for (5.7) and its restatement, (A5). In the slightly modified form given below, it also provides a counter-example for (A6). If we set

$$\varphi(z) = z_1^2 - z_1z_2 - 2z_1z_3 + z_2^2 + z_2z_3 + z_3^2$$

and consider the point $z = (1, 1, 0)$, then we have

$$\varphi(z) = 1, \varphi^{min} = 0$$

$$\Delta_1 = \Delta_2 = \Delta_3 = 1/4$$

and the three assertions mentioned previously are violated.

As things stand, convergence to φ^{min} cannot be guaranteed for the procedure described in the article if A is semi-definite. However, a possible solution to Problem I obtained by the procedure can readily be checked and such solutions have proved valid in the several applications I have seen. The simplicity of the procedure makes me believe that it would be useful to find conditions for the semi-definite case under which convergence to a minimum is assured.

I regret any confusion that my error may have caused.

signed Clifford Hildreth
Michigan State University

NEWS AND MEMORANDA

Readers are invited to submit to the Managing Editor items of general interest in the field of logistics.

Appreciation is expressed to the following for their valuable services as Associate Editors of this journal during their terms of office: J. D. Hayes, Rear Admiral, USN (retired), T. C. Roberts, Captain, SC, USN, M. A. Woodbury, New York University.

Vice Admiral R. E. McShane, USN (retired) and Dr. J. D. Wilkes have been appointed Associate Editors.

The following information has been furnished by the Industrial College of the Armed Forces, Washington, D. C.:

The study of economic mobilization, a subject of paramount importance to the security of this Nation, demands the thoughtful consideration of military and civilian leaders during this period of international tension. It is the mission of the Industrial College of the Armed Forces, operating under the direct supervision of the Joint Chiefs of Staff, to provide responsible and interested individuals with a thorough understanding of the economic aspects of national security. The College concentrates on the procedures and problems involved in maintaining national readiness and, if the occasion demands, converting our economy to wartime needs.

Thousands of military officers and civilian leaders have taken advantage of the educational opportunities afforded by the Industrial College of the Armed Forces through its Extension Division, which offers two separate courses throughout the country. These courses are based on the Resident Course of the College, which is offered to 145 senior military officers and top-ranking civilians from Federal agencies each year.

The first of these extension programs is a correspondence course entitled "Emergency Management of the National Economy," which is available to field grade officers, both Regulars and Reserves, as well as to civilians of comparable background and experience. This course is designed to reach key personnel on both sides of the important civilian-military team, which is vital to national preparedness. The many facets of mobilization, each of which has an important influence on the economic might of the United States, are indicated by a review of the

titles of the 22 textbooks in the course. All graduates of the course may retain these books, which make a valuable addition to any personal library.

Volume	Title
I.	The Nature of Economic Mobilization
II.	Basic Economics — Refresher Notes
III.	Public Support, Morale and Security
IV.	Principles of Administration
V.	Human Resources
VI.	Natural Resources
VII.	Energy Resources
VIII.	Research and Development
IX.	Transportation
X.	Public Utilities and Services
XI.	Requirements
XII.	Procurement
XIII.	Production
XIV.	Military Supply Management
XV.	Economic Warfare and Economic Intelligence
XVI.	U.S. Economic Foreign Policy
XVII.	Foreign Aid
XVIII.	Psychological Aspects of Global Conflict
XIX.	Administration of Mobilization WW II
XX.	Economic Stabilization
XXI.	Reconversion and Partial Mobilization
XXII.	Retrospect and Prospect

Many persons, both military and civilian, have been connected with a single segment of a mobilization effort, during either World War II or the Korean conflict, but few have had an opportunity to observe directly the interrelationship of the various aspects of economic readiness. The Correspondence Course draws together and explains the many elements of mobilization, giving the students an overall description of the subject. The course does not attempt to "sell" any theories or expound tailored solutions to the problems involved in mobilization. Rather, the course invites the students to think for themselves by offering an analysis of previous mobilization efforts and presenting the current thoughts of economic and industrial experts.

The Industrial College awards a certificate of completion to students who satisfactorily complete this course, "Emergency Management of the National Economy." Completion of the course is also recorded on the official personnel records of all military officers, and as desired by civilian students. This accomplishment is thus readily available for consideration by anyone who may review the graduate's record in an official capacity. Officers of the Reserve

or National Guard organizations who are not on extended active duty are also accorded a total of 48 credit points for promotion and retirement purposes upon satisfactory completion of the entire course.

More than 5,500 students have graduated from the Correspondence Course to date. Many of these graduates, both military and civilian, have indicated that the course enables them to perform more effectively in their fields of endeavor. These graduates are better informed and more responsible citizens who during this "age of peril" contribute to the security of the country.

Either military or civilian personnel desirous of taking this course may communicate directly with the Correspondence Branch, Industrial College of the Armed Forces, Washington 25, D. C. A prospectus that outlines eligibility requirements will be forwarded together with an application for enrollment.

The Extension Division of the Industrial College also reaches qualified individuals throughout the country with a series of National Resources Conferences, which are held in sixteen cities each year. Each Conference consists of a series of lectures by the faculty on the same subjects taught by the Resident and Correspondence Courses, followed by analytical discussions by both the faculty and the conferees. The program includes discussions of such topics as the impact of technological progress on the national economy, the history and political development of Soviet Communism, the present status of the Nation's natural and human resources, and the wartime economic potential of other countries.

Teams of Army, Navy, Air Force, and Marine officers from the faculty of the Industrial College present the two-week conferences in each city. Each conference schedule involves four hours a day, five days a week, and consists of approximately 30 illustrated lectures, group discussions, a few movies, and usually a trip to a local industrial plant. During 1958, National Resources Conferences will be held in the following cities:

Fresno, Calif.	13 Jan-24 Jan 58
Indianapolis, Ind.	13 Jan-24 Jan 58
Los Angeles, Calif.	10 Feb-21 Feb 58
Boston, Mass.	10 Feb-21 Feb 58
Charleston, W. Va.	17 Mar-28 Mar 58
Philadelphia, Pa.	17 Mar-28 Mar 58
Rochester, N. Y.	14 Apr-25 Apr 58
Fort Worth, Tex.	14 Apr-25 Apr 58
Montgomery, Ala.	12 May-23 May 58
Bridgeport, Conn.	12 May-23 May 58

So far the conferences have reached over 38,000 officers and civilians. Civilian conferees are chosen by a civilian selection committee in each city. They are selected as a representative cross section of industry, agriculture, labor, business, the professions, religion, education, women's organizations, and civic community life.

Army, Air Force, and Naval Reserve officers, not on extended active duty who desire to attend the conferences may apply through official channels to their respective Army area and Air Force Commanders, or Naval District Commandants. Marine and Coast Guard Reserve officers apply respectively to Headquarters, Marine Corps, and Headquarters, U.S. Coast Guard. Upon selection, Reserve officers are ordered to active duty for the two-week period of instruction.

Regular officers and Reserve and National Guard officers on extended active duty may apply for attendance orders from their immediate headquarters following approval by the conference administrator.

All persons who attend a required percentage of lectures, films, and discussion periods are presented with a certificate of completion. In addition, military personnel receive active duty training credits provided for by pertinent regulations.

Both the Correspondence Course and the National Resources Conferences provide the people of the United States with an understanding of the plans and problems involved in maintaining our national security. These Extension Courses of the Industrial College have reached over 43,000 citizens who are now better qualified to administer and protect our heritage of freedom. By taking advantage of the unique opportunities afforded by the Industrial College of the Armed Forces, qualified individuals are contributing to their own education and adding to the security of the United States.

It has been announced by the Department of Commerce that:

A mathematical discussion of the theory of replacement of machinery with random failure rate is contained in a report of Army research just released for industry use through the Office of Technical Services, U.S. Department of Commerce.

The study analyzes the effect of recurrent replacements on the expected time-to-failure of individual components or groups of components. The recurrent replacements are of two types, strictly periodic and random. Random replacement has application where the work cycle of a given device is variable, and the device cannot be replaced in mid-cycle. Under periodic replacement is included the case of an infinite period, or replacement on failure, strictly periodic replacement, and replacement with a random period.

Rational bases are derived for formulation of a replacement policy. Expressions are given for the reliability function with replacement in terms of that without replacement.

The report, PB 121424 On the Theory of Replacement of Machinery with Random Failure Time, G. H. Weiss, Aberdeen Proving Ground, U.S. Army Ordnance, March 1956, may be ordered from OTS, U.S. Department of Commerce, Washington 25. It contains 25 pages, 75 cents.

* * *

RECENT PUBLICATIONS

INTRODUCTION TO FINITE MATHEMATICS. By John G. Kemeny, J. Laurie Snell, and Gerald L. Thompson. Prentice Hall, New York, 1957. 372 + xi pp. $5.00.

This book cannot be recommended highly enough to behavioral scientists who like to master the type of mathematics that will, to an ever increasing degree, be applied to their fields. Also, undergraduates in mathematics who like to expand their knowledge beyond the traditional calculus will find in this book the foundations of a type of mathematics that, in various forms, is beginning to be applied to many branches of science.

This book starts off with the logic, symbolized by the configuration of a tree, underlying this modern mathematics. This is followed by a clear exposition of the basic operations on sets and the principles of partitioning and counting. One chapter deals specifically with probabilities, culminating in a lucid description of Markov chains. These lead almost naturally to a discussion of vectors and matrices, with interesting applications to Markov chains and permutations. Then follows a brief exposition of linear programming and the theory of games, two recent branches of mathematics that have proved to be useful in complicated situations. The last chapter deals with fascinating applications to sociology (dominance relationships and communication networks), to genetics (stochastic processes), to psychology (Estes learning model), to anthropology (choice of marriage rules), and to economics (equilibrium of an expanding economy).

Particularly, these applications show how useful finite mathematics can be for the solution of a broad class of problems facing the behavioral scientist of our day. At the same time, it shows the development of mathematics as a result of the search of the mathematician to answer the questions raised in other branches of science. All those who take a lively interest in the development of mathematics should therefore not miss this excellent introduction to a new field that will prove to be of increasing importance in the years to come.

E. H. Bowman and R. B. Fetter, Analysis for Production Management, Homewood, Illinois: Richard Irwin, 1957. 503 + xiii pp., $8.40.

This is a textbook, complete with problems at the ends of the sections and ten cases for analysis. As such it fills an important gap in the previously available literature.

The first two chapters constitute an introduction to "scientific methods" in management. Chapter 3 takes up schematic models (such as flow charts). Chapters 4 and 5 discuss linear programming problems and calculations, and touch upon dynamic programming. Statistical Control, Sampling Inspection, and Industrial Experimentation are briefly covered by Chapters 6, 7, and 8. Chapters 9, 10, 11, and 12 take up Total Value Analysis, Incremental Analysis, Monte Carlo Analysis, and Equipment Investment Analysis. All of these topics are covered largely through examples. An epilogue, tying the general ideas of the book together, follows Chapter 12, then the ten cases, and finally a few appropriate statistical tables.

The reviewer has used this book as the principal text for a business-school seminar in "advanced production control." It turned out to be teachable to students with only minimum mathematical and statistical backgrounds, although (as is hardly surprising) some of the sections required a "cookbook" approach for those who had not had courses in calculus. Especially impressive was the extent to which many problems were designed so that students were effectively guided in going well beyond the material explicitly covered by the text. The cases, unfortunately, often degenerate into almost-disconnected sequences of obvious problems, laid end to end, but nevertheless proved to be valuable teaching aids.

In addition to its utility as a classroom text, the reviewer believes that Analysis for Production Management has a significant place as book for independent study by mature, but not necessarily mathematically sophisticated readers who wish to gain insight concerning the growing application of mathematical methods to management problems, and even the beginnings of real working knowledge. And, while this book is oriented toward production management, its central ideas are equally relevant to the more general topic of "analysis for management in general." Whether viewed broadly or narrowly, it is a welcome addition to management literature.

* * *

AUTHOR INDEX FOR VOLUME IV

	Pages
BARTLETT, T. E., "An Algorithm for the Minimum Number of Transportation Units Required to Maintain a Specified Schedule"	139 - 149
BARTLETT, T. E., and CHARNES, A., "Cyclic Scheduling and Combinatorial Topology; Assignment and Routing of Motive Power to Meet Scheduling and Maintenance Requirements." Part II, "Generalization and Analysis"	207 - 220
BELLMAN, Richard, "Formulation of Recurrence Equations for Shuttle Process and Assembly Line"	321 - 334
BREMER, Hans, HALL, William, and PAULSEN, Maxine, "Experiences with the Bid Evaluation Problem" (Abstract)	27
CAMPBELL, R. D., DOREY, F. D., and MURPHY, R. E., "Concept of a Logistics System"	101 - 116
CHARNES, A., and BARTLETT, T. E., "Cyclic Scheduling and Combinatorial Topology; Assignment and Routing of Motive Power to Meet Scheduling and Maintenance Requirements." Part II, "Generalization and Analysis"	207 - 220
CHARNES, A., and MILLER, M. H., "Mathematical Programming and Evaluation of Freight Shipment Systems, Application and Analysis." Part II, "Analysis"	243 - 252
CHENEV, L. K., "Linear Program Planning of Refinery Operations"	9 - 16
CRANE, Roger R., "Some Recent Developments in Transportation Research"	173 - 181
DOREY, F. D., CAMPBELL, R. D., and MURPHY, R. E., "Concept of a Logistics System"	101 - 116
DWYER, Paul S., and GALLER, Bernard, "Translating the Method of Reduced Matrices to Machines"	55 - 71
FINN, W. R., and WILSON, A. H., "Improvise or Plan?"	263 - 267
FORD, L. R., and FULKERSON, D. R., "A Primal-Dual Algorithm for the Capacitated Hitchcock Problem"	47 - 54
FRIEDMAN, Lawrence, "An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline." Annex B, "Calculation of the School Table"	199 - 202

	Pages
FRIEDMAN, Lawrence, and YASPAN, Arthur J., "An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline." Annex A, "The Assignment Problem Technique"	193 - 197
FULKERSON, D. R., and FORD, L. R., Jr., "A Primal-Dual Algorithm for the Capacitated Hitchcock Problem"	47 - 54
GALLER, Bernard, and DWYER, Paul S., "Translating the Method of Reduced Matrices to Machines"	55 - 71
GASS, Saul I., "On The Distribution of Manhours to Meet Scheduled Requirements"	17 - 25
GEISLER, Murray A., "Relationships Between Weapons and Logistics Expenditures"	335 - 346
GLEAVES, V. B., "Cyclic Scheduling and Combinatorial Topology; Assignment and Routing of Motive Power to Meet Scheduling and Maintenance Requirements." Part I, "A Statement of the Operating Problem of the Frisco Railroad"	203 - 205
HALE, J. K., and WICKE, H. H., "An Application of Game Theory to Special Weapons Evaluation"	347 - 356
HALL, William, PAULSEN, Maxine, and BREMMER, Hans, "Experiences with the Bid Evaluation Problem" (Abstract)	27
HANSSMAN, F., and McCLOSKEY, J. F., "An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline"	183 - 191
HELLER, I., "Constraint Matrices of Transportation-Type Problems" (Abstract)	73 - 76
HENN, Carl L., LCDR, "Multinational Logistics in the Nuclear Age"	117 - 130
HILDRETH, Clifford, "A Quadratic Programming Procedure"	79 - 85
ISBELL, J. R., "An Optimal Search Pattern"	357 - 359
JACKSON, J. R., "Simulation Research on Job Shop Production"	287 - 295
JACOBS, Walter, "Loss of Accuracy in Simplex Computations"	89 - 94
JEWELL, William S., "Warehousing and Distribution of a Seasonal Product"	29 - 34
JUNCOSA, M. L., and KALABA, R. E., "Communication-Transportation Networks"	221 - 222
KALABA, R. E., and JUNCOSA, M. L., "Communication-Transportation Networks" (Abstract)	221 - 222
KARUSH, William, and VAZSONYI, A., "Mathematical Programming and Employment Scheduling"	297 - 320
KELLEY, J. E., Jr., "A Threshold Method for Linear Programming"	35 - 45

	Pages
KLEIN, Morton, "Some Production Planning Problems"	269 - 286
MARKOWITZ, Harry, "Computing Procedures for Portfolio Selection" (Abstract)	87 - 88
McCLOSKEY, J. F., and HANSSMAN, F., "An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline"	183 - 191
MILLER, M. H., and CHARNES, A., "Mathematical Programming and Evaluation of Freight Shipment Systems, Application and Analysis." Part II, "Analysis"	243 - 252
MUNKRES, James, "On the Assignment and Transportation Problems" (Abstract)	77 - 78
MURPHY, R. E., DOREY, F. D., and CAMPBELL, R. D., "Concept of a Logistics System"	101 - 116
O'NEILL, R. E., "Analysis and Monte Carlo Simulation of Cargo Handling"	223 - 236
PAULSEN, Maxine, HALL, William, and BREMER, Hans, Experiences with the Bid Evaluation Problem" (Abstract)	27
PRAGER, William, "Numerical Solution of the Generalized Transportation Problem"	253 - 261
ROBBINS, James J., "Tank Duel with Game Theory Implications" (translation of a paper by L. E. Zachrisson)	131 - 138
ROSENBLATT, David, "On the Graphs and Asymptotic Forms of Finite Boolean Relation Matrices and Stochastic Matrices"	151 - 167
SOYSTER, H. R., "Mathematical Programming and Evaluation of Freight Shipment Systems, Application and Analysis." Part I, "Applications"	237 - 242
SUZUKI, George, "Procurement and Allocation of Naval Electronic Equipments"	1 - 7
THOMPSON, P. M., "Editing Large Linear Programming Matrices"	97 - 100
TOMPKINS, C., "Some Methods of Computational Attack on Programming Problems Other than the Simplex Method" (Abstract)	95 - 96
VAZSONYI, A., and KARUSH, William, "Mathematical Programming and Employment Scheduling"	297 - 320
WICKE, H. H., HALE, J. K., "An Application of Game Theory to Special Weapons Evaluation"	347 - 356
WILSON, A. H., and FINN, W. R., "Improvise or Plan?"	263 - 267

Pages

YASPAN, Arthur J., and FRIEDMAN, Lawrence, Annex A, "The Assignment Problem Technique," of the paper "An Analysis of Stewardess Requirements and Scheduling for a Major Domestic Airline" . . . . . 193 - 197

ZACHRISSON, L. E., "Tank Duel with Game Theory Implications" (translated by J. J. Robbins) . . . . . 131 - 138

* * *

☆ U. S. GOVERNMENT PRINTING OFFICE : 1957 O - 449503

22688 152